

Stochastic Recursive Inclusions with Non-Additive Iterate-Dependent Markov Noise

Vinayaka G. Yaji and Shalabh Bhatnagar

July 19, 2016

Abstract

In this paper we study the asymptotic behavior of stochastic approximation schemes with set-valued drift function and non-additive iterate-dependent Markov noise. We show that a linearly interpolated trajectory of such a recursion is an asymptotic pseudotrajectory for the flow of a limiting differential inclusion obtained by averaging the set-valued drift function of the recursion w.r.t. the stationary distributions of the Markov noise. The limit set theorem in [1] is then used to characterize the limit sets of the recursion in terms of the dynamics of the limiting differential inclusion. We then state two variants of the Markov noise assumption under which the analysis of the recursion is similar to the one presented in this paper. Scenarios where our recursion naturally appears are presented as applications. These include controlled stochastic approximation, subgradient descent, approximate drift problem and analysis of discontinuous dynamics all in the presence of non-additive iterate-dependent Markov noise.

1 Introduction

Consider the standard stochastic approximation scheme given by,

$$X_{n+1} = X_n + a(n) (h(X_n) + M_{n+1}) \quad (1)$$

where $\{X_n\}_{n \geq 0}$ is a sequence of \mathbb{R}^d -valued random variables, $h : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a Lipschitz continuous drift function and $\{M_n\}_{n \geq 1}$ is a sequence of \mathbb{R}^d -valued random variables which denote the additive noise terms. In [2], it was shown that under certain assumptions the asymptotic behavior of recursion (1) can be determined by the asymptotic behavior of the *o.d.e.*

$$\frac{dx}{dt} = h(x).$$

This method is known as the ODE method and the central idea of this method is to show that the linearly interpolated trajectory of recursion (1) “tracks” the flow of the o.d.e. This idea was later generalized in [3] to analyze the asymptotic behavior of stochastic processes with continuous sample paths for which a tracking argument as in the ODE method could be established. It was here that the notion of *asymptotic pseudotrajectory* was introduced. For a precise definition we refer the reader to chapter 3 of [3]. In many applications arising in machine learning and optimization, the drift function h is set-valued or is single-valued and does not satisfy the assumption of Lipschitz continuity (or even continuity). In such cases the recursion studied takes the form,

$$X_{n+1} - X_n - a(n)M_{n+1} \in a(n)H(X_n), \quad (2)$$

where $H : \mathbb{R}^d \rightarrow \{\text{subsets of } \mathbb{R}^d\}$ is a set-valued map satisfying certain assumptions. Such recursions were first studied in [1] by showing that the linearly interpolated trajectory of recursion (2) “tracks” the flow of the differential inclusion given by,

$$\frac{dx}{dt} \in H(x).$$

The notion of asymptotic pseudotrajectory was also extended to the set-valued case and a limit set theorem established.

In many applications such as methods relying on Monte Carlo simulation where sampling exactly from a distribution (or family of distributions) is not possible and instead Markov chain Monte Carlo methods are used, the recursion consists of an additional non-additive iterate-dependent Markov noise component. The recursion takes the general form,

$$X_{n+1} - X_n = a(n) (h(X_n, S_n) + M_{n+1}), \quad (3)$$

where $\{S_n\}_{n \geq 0}$ denotes the Markov noise component. The other quantities in the above recursion have the same interpretation as those in (1). The analysis of such a recursion was first performed in [4]. In [4], the analysis is carried out under the assumptions that the drift function h is continuous, the Markov chain defined by the transition kernel associated with the Markov noise terms in the recursion admits a unique stationary distribution as well as a solution to the Poisson equation among others. The case with discontinuities in the drift function was considered in [5] but the analysis required the function $h(\cdot, s)$ to be Holder continuous. Recently in [6], such a recursion has been analyzed when the drift function is just measurable. In all the above references, however, the Markov noise terms are assumed to have a unique stationary distribution and the assumptions imply an existence of a solution to the Poisson equation. Another set of assumptions on the Markov noise terms under which the recursion (3) is studied can be found in [7]. In [7], the transition kernel defining the Markov noise terms is required to be continuous in both the iterate and the state variables. The advantage in this case is that the set of stationary distributions need not be unique and one does not need the Markov chain associated with the transition kernel for every value of the iterate, to be aperiodic and irreducible. Further the analysis in [7] enables one to study the recursion (3) when the noise terms are not Markov by themselves, but their lack of Markov property comes through the dependence on a control sequence.

In this paper we extend the above to the case where the drift function h in recursion (3), is a set-valued map. The recursion now takes the form,

$$X_{n+1} - X_n - a(n)M_{n+1} \in a(n)H(X_n, S_n),$$

where $H : \mathbb{R}^d \times \mathcal{S} \rightarrow \{\text{subsets of } \mathbb{R}^d\}$ is a set-valued map and \mathcal{S} denotes the state space of the Markov noise terms which we assume to be a compact metric space. The assumption on the Markov noise terms is similar to [7]. We shall show that the linearly interpolated trajectory of the above recursion is an asymptotic pseudotrajectory for the flow of a limiting differential inclusion obtained by averaging the set-valued map w.r.t. the stationary distributions of the Markov noise terms. The main idea is to approximate the set-valued drift function with continuous set-valued maps which admit a single-valued parametrization, thereby enabling us to write the recursion as a standard stochastic approximation scheme with single-valued maps for which an asymptotic pseudotrajectory argument is easy to establish. Later we invoke the limit set theorem in [1] to characterize the limit set of the above recursion.

Organization of the paper:

In section 2, we formally define the recursion and state the assumptions imposed, which is then followed by a brief discussion on each of the assumptions. In section 3, we review certain results from set-valued analysis which will be used later to analyze the recursion and define the limiting differential inclusion. In section 4, the limiting differential inclusion is defined and properties of the same are established. In section 5, we define the space of probability measure valued functions and define an appropriate topology on this space. A metrization lemma for the above mentioned space is stated and a continuous function on such a space needed later in analyzing the recursion is defined. In section 6, we first state some preliminaries and later state our main result followed by the limit set theorem. In section 7, we state two variants of the Markov noise assumption under which the analysis of the stochastic approximation scheme can be carried out along similar lines. In section 8, we state four applications where the recursion

analyzed in this paper naturally appears. We finally conclude in section 9 by providing a few interesting directions for future work.

2 Recursion and assumptions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{X_n\}_{n \geq 0}$ be a sequence of \mathbb{R}^d -valued random variables on Ω satisfying

$$X_{n+1} - X_n - a(n)M_{n+1} \in a(n)H(X_n, S_n) \quad (4)$$

where,

(A1) $H : \mathbb{R}^d \times \mathcal{S} \rightarrow \{\text{subsets of } \mathbb{R}^d\}$ is a set valued map on $\mathbb{R}^d \times \mathcal{S}$ where $(\mathcal{S}, d_{\mathcal{S}})$ is a compact metric space and the map $H(\cdot)$ satisfies

- (i) for every $(x, s) \in \mathbb{R}^d \times \mathcal{S}$, $H(x, s)$ is a non-empty, compact, convex subset of \mathbb{R}^d ,
- (ii) there exists $K > 0$ such that, for every $(x, s) \in \mathbb{R}^d \times \mathcal{S}$,
 $\sup_{z \in H(x, s)} \|z\| \leq K(1 + \|x\|)$,
- (iii) for every $\mathbb{R}^d \times \mathcal{S}$ valued sequence, $\{(x_n, s_n)\}_{n \geq 1}$ converging to (x, s) , for every sequence $\{z_n \in H(x_n, s_n)\}_{n \geq 1}$ converging to z , we have, $z \in H(x, s)$.

Throughout this paper $\mathcal{P}(\cdots)$, denotes the set of probability measures on a compact metric space ‘ \cdots ’ with the Prohorov topology (see chapter 2 in [8]).

(A2) $\{S_n\}_{n \geq 0}$ is a sequence of \mathcal{S} valued random variables on Ω such that, for every $n \geq 0$, for every $A \in \mathcal{B}(\mathcal{S})$, $\mathbb{P}(S_{n+1} \in A | S_n, X_n) = \mathbb{P}(S_{n+1} \in A | S_n, X_n) = \Pi(X_n, S_n)(A)$ a.s. where $\Pi : \mathbb{R}^d \times \mathcal{S} \rightarrow \mathcal{P}(\mathcal{S})$, is continuous.

(A3) $\{a(n)\}_{n \geq 0}$ is a sequence of real numbers satisfying

- (i) $a(0) \leq 1$ and for every $n \geq 0$, $a(n) \geq a(n+1)$,
- (ii) $\sum_{n \geq 0} a(n) = \infty$ and $\sum_{n \geq 0} a(n)^2 < \infty$.

(A4) $\{M_n\}_{n \geq 1}$ is a sequence of \mathbb{R}^d valued random variables such that, a.e. (ω) , for every $T > 0$

$$\lim_{n \rightarrow \infty} \sup_{n \leq k \leq \tau(n, T)} \left\| \sum_{m=n}^k a(m)M_{m+1} \right\| = 0 \text{ where, } \tau(n, T) := \min \left\{ m > n : \sum_{k=n}^{m-1} a(k) \geq T \right\}.$$

(A5) $\mathbb{P}(\sup_{n \geq 0} \|X_n\| < \infty) = 1$.

Assumption (A1) is an extension of the assumption imposed on the set valued map for the case without Markov noise in [1] to the case with Markov noise. The only difference is in (A1)(ii), since in our case we expect the constant K to be independent of Markov noise terms. This strengthening of the assumption allows us to obtain integrable measurable selections and also to prove certain properties of the limiting differential inclusion in future sections.

Assumption (A2) ensures that for every $x \in \mathbb{R}^d$ the Markov chain defined by the transition kernel $\Pi(x, \cdot)(\cdot)$ possesses weak Feller property (see chapter 6.1.1 of [9]). In addition to the above, since the state space of the Markov chain is a compact metric space, by Theorem 12.1.2(ii) in [9], for every $x \in \mathbb{R}^d$, the Markov chain defined by the transition kernel $\Pi(x, \cdot)(\cdot)$ admits at least one stationary distribution ($\mu \in \mathcal{P}(\mathcal{S})$ is stationary for the Markov chain defined by the transition kernel $\Pi(x, \cdot)(\cdot)$ if, for every $A \in \mathcal{B}(\mathcal{S})$, $\mu(A) = \int_{\mathcal{S}} \Pi(x, s)(A) \mu(ds)$). For every $x \in \mathbb{R}^d$, let $D(x) \subseteq \mathcal{P}(\mathcal{S})$ denote the set of stationary distributions of the Markov chain defined by the transition kernel $\Pi(x, \cdot)(\cdot)$. It can be easily shown that,

- (i) for every $x \in \mathbb{R}^d$, $D(x)$ is a convex and compact subset of $\mathcal{P}(\mathcal{S})$.

(ii) the map $x \rightarrow D(x)$ has a closed graph, that is the set

$$\left\{ (x, \mu) \in \mathbb{R}^d \times \mathcal{P}(\mathcal{S}) : x \in \mathbb{R}^d, \mu \in D(x) \right\},$$

is a closed subset of $\mathbb{R}^d \times \mathcal{P}(\mathcal{S})$ (for a proof of the above two properties we refer the reader to page 69 in [10]).

It is worth noting here that for every $x \in \mathbb{R}^d$, we do *not* require that the associated Markov chain be aperiodic, irreducible or possess a unique stationary distribution.

Assumption (A3) is the standard step size assumption. The assumptions of a non-increasing sequence and square summability are required in the analysis of the Markov noise component.

Assumption (A4) is the same additive noise assumption as in Proposition 1.3, section 1.5 of [1]. The assumption ensures that the contribution of the additive noise is eventually negligible. For various sufficient conditions under which (A4) is satisfied we refer the reader to [1].

Assumption (A5) ensures that the sample path of iterates of recursion (4) remain bounded. This assumption is known as the stability assumption and sufficient conditions for ensuring the same for the single-valued case (see [11], [12]) and for the set-valued case without Markov noise (see [13]) exist.

3 Background

In this section we shall review some results from the theory of set valued maps and their integration. First we shall present a result on approximation of upper semi-continuous set valued maps with continuous set valued maps. We will then use this approximation to obtain single valued maps that represent these approximate continuous set valued maps thus enabling us to rewrite the stochastic recursive inclusion, (4), as a standard stochastic approximation scheme with single valued maps.

Secondly, in order to define the set valued map associated with the limiting differential inclusion (also known as mean field) and prove properties of the same we review definitions of measurability of set valued maps, its integral and some properties of the integral.

3.1 Upper semicontinuous set valued maps and their approximation

First we shall recall the notion of upper semicontinuity, lower semicontinuity and continuity of set valued maps. These notions are taken from chapter 1, section 1 of [14].

Definition A set valued map $H : \mathbb{R}^d \times \mathcal{S} \rightarrow \{\text{subsets of } \mathbb{R}^d\}$ is,

- *Upper semicontinuous* (u.s.c.) if, for every $(x_0, s_0) \in \mathbb{R}^d \times \mathcal{S}$, for every $\epsilon > 0$, there exists $\delta > 0$ (depending on (x_0, s_0) and ϵ) such that,

$$\|x - x_0\| < \delta, d_{\mathcal{S}}(s, s_0) < \delta \implies H(x, s) \subseteq H(x_0, s_0) + \epsilon U,$$

where U denotes the closed unit ball in \mathbb{R}^d .

- *Lower semicontinuous* (l.s.c) if, for every $(x_0, s_0) \in \mathbb{R}^d \times \mathcal{S}$, for every $z_0 \in H(x_0, s_0)$, for every sequence $\{(x_n, s_n)\}_{n \geq 1}$ converging to (x_0, s_0) , there exists a sequence $\{z_n \in H(x_n, s_n)\}$ converging to z_0 .
- *Continuous* if, it is both u.s.c. and l.s.c.

For set valued maps taking compact set values we have the above mentioned notion of u.s.c. to be equivalent to the standard notion of u.s.c. (see pg. 45, [14]). In this paper we shall encounter set valued maps which are compact set valued and hence we have chosen to state the above as the definition of upper semicontinuity.

Since the set valued map H satisfying assumption (A1) has closed graph (i.e. assumption (A1)(iii) holds), the following lemma follows from Corollary 1 in chapter 1, section 1 of [14].

Lemma 3.1 [u.s.c.] *A set valued map H satisfying (A1) is u.s.c.*

The next lemma gives a sequence of decreasing continuous set valued maps which approximate the set valued map H satisfying assumption (A1). The statement of the lemma below can be found in page 39 of [14].

Lemma 3.2 [continuous embedding] *Let H be a set valued map satisfying (A1). Then there exists a sequence of set valued maps, $\{H^{(l)}\}_{l \geq 1}$ where for every $l \geq 1$,*

- (i) $H^{(l)} : \mathbb{R}^d \times \mathcal{S} \rightarrow \{\text{subsets of } \mathbb{R}^d\}$ is continuous and for every $(x, s) \in \mathbb{R}^d \times \mathcal{S}$, $H^{(l)}(x, s)$ is a convex and compact subset of \mathbb{R}^d ,
- (ii) for every $(x, s) \in \mathbb{R}^d \times \mathcal{S}$, $H(x, s) \subseteq H^{(l+1)}(x, s) \subseteq H^{(l)}(x, s)$,
- (iii) there exists $K^{(l)} > 0$, such that for every $(x, s) \in \mathbb{R}^d \times \mathcal{S}$,
 $\sup_{z \in H^{(l)}(x, s)} \|z\| \leq K^{(l)}(1 + \|x\|)$.

Furthermore, for every $(x, s) \in \mathbb{R}^d \times \mathcal{S}$, $H(x, s) = \bigcap_{l \geq 1} H^{(l)}(x, s)$.

A brief outline of the proof of the above lemma is provided in appendix A. The following are some useful observations from the proof of Lemma 3.2.

- (a) $\tilde{K} := \sup_{l \geq 1} K^{(l)}$ is finite.
- (b) For every $(x, s) \in \mathbb{R}^d \times \mathcal{S}$, for every $\epsilon > 0$, there exists L (depending on ϵ and (x, s)), such that, for every $l \geq L$, $H^{(l)}(x, s) \subseteq H(x, s) + \epsilon U$.

The next lemma gives us a continuous parametrization of the continuous set valued maps $H^{(l)}$ as in Lemma 3.2. This lemma follows from Theorem 2 in chapter 1, section 7 of [14].

Lemma 3.3 [parametrization] *For every $l \geq 1$, $H^{(l)}$ be a set valued map as in Lemma 3.2. Then for every $l \geq 1$, there exists a continuous function $h^{(l)} : \mathbb{R}^d \times \mathcal{S} \times U \rightarrow \mathbb{R}^d$ such that,*

- (i) for every $(x, s) \in \mathbb{R}^d \times \mathcal{S}$, $H^{(l)}(x, s) = h^{(l)}(x, s, U)$ where U denotes the closed unit ball in \mathbb{R}^d and $h^{(l)}(x, s, U) := \{h^{(l)}(x, s, u) : u \in U\}$,
- (ii) for every $(x, s, u) \in \mathbb{R}^d \times \mathcal{S} \times U$, $\|h^{(l)}(x, s, u)\| \leq K^{(l)}(1 + \|x\|)$ where $K^{(l)}$ is as in Lemma 3.2(iii).

Combining Lemma 3.2 and Lemma 3.3 we obtain the approximation theorem stated below.

Theorem 3.4 [approximation] *Let H be a set valued map satisfying (A1). Then there exists a sequence of continuous functions, $\{h^{(l)}\}_{l \geq 1}$, such that, for every $l \geq 1$,*

- (i) $h^{(l)} : \mathbb{R}^d \times \mathcal{S} \times U \rightarrow \mathbb{R}^d$ is continuous and for every $(x, s) \in \mathbb{R}^d \times \mathcal{S}$, $h^{(l)}(x, s, U)$ is a convex and compact subset of \mathbb{R}^d ,
- (ii) for every $(x, s) \in \mathbb{R}^d \times \mathcal{S}$, $H(x, s) \subseteq h^{(l+1)}(x, s, U) \subseteq h^{(l)}(x, s, U)$,
- (iii) there exists $K^{(l)} > 0$, such that for every $(x, s, u) \in \mathbb{R}^d \times \mathcal{S} \times U$,
 $\|h^{(l)}(x, s, u)\| \leq K^{(l)}(1 + \|x\|)$.

Furthermore, for every $(x, s) \in \mathbb{R}^d \times \mathcal{S}$, $H(x, s) = \bigcap_{l \geq 1} h^{(l)}(x, s, U)$.

3.2 Measurable set valued maps and integration

Let $(\mathcal{Y}, \mathcal{F}_{\mathcal{Y}})$ denote a measurable space and $F : \mathcal{Y} \rightarrow \{\text{subsets of } \mathbb{R}^d\}$ be a set valued map such that, for every $y \in \mathcal{Y}$, $F(y)$ is a non-empty closed subset of \mathbb{R}^d . Throughout this section F refers to the set valued map as defined above.

Definition [*measurable set-valued map*] A set valued map F is measurable if for every $C \subseteq \mathbb{R}^d$, closed,

$$F^{-1}(C) := \{y \in \mathcal{Y} : F(y) \cap C \neq \emptyset\} \in \mathcal{F}_{\mathcal{Y}}.$$

We refer the reader to Theorem 1.2.3 in [15] for other notions of measurability and their relation to the definition above.

Definition [*measurable selection*] A function $f : \mathcal{Y} \rightarrow \mathbb{R}^d$ is a measurable selection of a set valued map F if f is measurable and for every $y \in \mathcal{Y}$, $f(y) \in F(y)$.

For a set valued map F let, $\mathcal{S}(F)$ denote the set of all measurable selections. The next lemma summarizes some standard results about measurable set valued maps and their measurable selections.

Lemma 3.5 *For any measurable set valued map F ,*

(i) $\mathcal{S}(F) \neq \emptyset$.

(ii) (*Castaing representation*) there exists $\{f_n\}_{n \geq 1} \subseteq \mathcal{S}(F)$ such that, for every $y \in \mathcal{Y}$, $F(y) = \text{cl}(\{f_n(y)\}_{n \geq 1})$, where $\text{cl}(\cdot)$ denotes the closure of a set.

We refer the reader to Theorem 1.2.6 and Theorem 1.2.7 in [15] for the proofs of Lemma 3.5(i) and (ii) respectively.

Definition [μ -integrable set-valued map] Let μ be a probability measure on $(\mathcal{Y}, \mathcal{F}_{\mathcal{Y}})$. A measurable set valued map F is said to be μ -integrable if, there exists $f \in \mathcal{S}(F)$ which is μ -integrable.

Definition [*Aumann's integral*] Let μ be a probability measure on $(\mathcal{Y}, \mathcal{F}_{\mathcal{Y}})$. The integral of an integrable set valued map F is defined as,

$$\int_{\mathcal{Y}} F(y) \mu(dy) := \left\{ \int_{\mathcal{Y}} f(y) \mu(dy) : f \in \mathcal{S}(F), f \text{ is } \mu\text{-integrable} \right\}.$$

The next lemma states a useful result on the properties of the integral of a set valued map which is convex and compact set valued.

Lemma 3.6 *Let μ be a probability measure on $(\mathcal{Y}, \mathcal{F}_{\mathcal{Y}})$ and F a μ -integrable set valued map such that, for every $y \in \mathcal{Y}$, $F(y)$ is convex and compact. Then, $\int_{\mathcal{Y}} F(y) \mu(dy)$ is a convex and closed subset of \mathbb{R}^d .*

For a proof of the above lemma we refer the reader to Theorem 2.2.2 in [15]. The next lemma is a useful tool to prove measurability of set valued maps that arise later in this paper.

Lemma 3.7 *Let $g : \mathcal{Y} \rightarrow \mathbb{R}^d$ be a measurable function and F a measurable set valued map. Then the map $y \rightarrow d(g(y), F(y))$ where $d(g(y), F(y)) := \inf_{z \in F(y)} \|g(y) - z\|$, is measurable.*

Proof Since F is measurable, there exists $\{f_n\}_{n \geq 1}$ a Castaing representation of F . Clearly for every $n \geq 1$, the map $y \rightarrow \|g(y) - f_n(y)\|$ is measurable. Hence, the map $y \rightarrow \inf_{n \geq 1} \|g(y) - f_n(y)\|$ is measurable. Since for every $y \in \mathcal{Y}$, $\text{cl}(\{f_n(y)\}_{n \geq 1}) = F(y)$ we have, $\inf_{n \geq 1} \|g(y) - f_n(y)\| = d(g(y), F(y))$. Therefore the map $y \rightarrow d(g(y), F(y))$ is measurable. ■

4 Limiting differential inclusion and its properties

In this section we shall define the limiting differential inclusion (or the mean field) whose flow the iterates of recursion (4) are expected to track. Analogous to the single valued case as in [7], the set valued map associated with the limiting differential inclusion is nothing but the set valued map obtained by averaging the set valued map H in equation (4) w.r.t. the stationary distributions of the Markov noise. The above shall be made precise in this section. Throughout this section we assume that H denotes the set valued map satisfying (A1) and $\{H^{(l)}\}_{l \geq 1}$, $\{h^{(l)}\}_{l \geq 1}$ are as in Lemma 3.2 and Lemma 3.3 respectively.

First we define slices of the set valued maps H and $H^{(l)}$ for every $x \in \mathbb{R}^d$.

Definition For every $x \in \mathbb{R}^d$, define,

- (i) $H_x : \mathcal{S} \rightarrow \{\text{subsets of } \mathbb{R}^d\}$ such that, for every $s \in \mathcal{S}$, $H_x(s) := H(x, s)$.
- (ii) for every $l \geq 1$, $H_x^{(l)} : \mathcal{S} \rightarrow \{\text{subsets of } \mathbb{R}^d\}$ such that, for every $s \in \mathcal{S}$, $H_x^{(l)}(s) := H^{(l)}(x, s)$.
- (iii) for every $l \geq 1$, $h_x^{(l)} : \mathcal{S} \times U \rightarrow \mathbb{R}^d$ such that, for every $(s, u) \in \mathcal{S} \times U$, $h_x^{(l)}(s, u) := h^{(l)}(x, s, u)$.

The lemma below summarizes properties that the set valued maps H_x and $H_x^{(l)}$ inherit from the set valued maps H and $H^{(l)}$ respectively.

Lemma 4.1 For every $x \in \mathbb{R}^d$,

- (i) H_x is a measurable set valued map and for every $s \in \mathcal{S}$, $H_x(s)$ is a convex and compact subset of \mathbb{R}^d . Further there exists $C_x := K(1 + \|x\|) > 0$, such that $\sup_{s \in \mathcal{S}} \sup_{z \in H_x(s)} \|z\| \leq C_x$ (which means H_x is bounded).
- (ii) for every $l \geq 1$, $H_x^{(l)}$ is a measurable set valued map and for every $s \in \mathcal{S}$, $H_x^{(l)}(s)$ is a convex and compact subset of \mathbb{R}^d . Further for every $l \geq 1$, there exists $C_x^{(l)} := K^{(l)}(1 + \|x\|) > 0$, such that $\sup_{s \in \mathcal{S}} \sup_{z \in H_x^{(l)}(s)} \|z\| \leq C_x^{(l)}$ (which means $H_x^{(l)}$ is bounded).
- (iii) For any probability measure μ on $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$, all measurable selections of H_x are μ -integrable and hence H_x is μ -integrable.
- (iv) For any probability measure μ on $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$, all measurable selections of $H_x^{(l)}$ are μ -integrable and hence $H_x^{(l)}$ is μ -integrable.
- (v) for every $l \geq 1$, $h_x^{(l)}$ is continuous, bounded by the constant $C_x^{(l)}$ where $C_x^{(l)}$ is as in part (ii) of this lemma and for every $s \in \mathcal{S}$, $h_x^{(l)}(s, U) = H_x^{(l)}(s)$.

Proof Fix some $x \in \mathbb{R}^d$.

- (i) For every $s \in \mathcal{S}$, the convexity and compactness of $H_x(s)$ follows from the definition of H_x and assumption (A1)(i). Boundedness of H_x follows from the definition of H_x and assumption (A1)(ii). From definition of H_x and assumption (A1)(iii), we deduce that for every sequence $\{s_n\}_{n \geq 1}$ converging to s and for every sequence $\{z_n \in H_x(s_n)\}_{n \geq 1}$ converging to z , $z \in H_x(s)$ (which means that the graph of H_x is closed).

Let C be some closed subset of \mathbb{R}^d . Consider a sequence $\{s_n\}_{n \geq 1}$ converging to s such that for every $n \geq 1$, $s_n \in \{s \in \mathcal{S} : H_x(s) \cap C \neq \emptyset\}$. For every $n \geq 1$, let $z_n \in H_x(s_n) \cap C$. Since H_x is bounded, $\{z_n\}_{n \geq 1}$ is a bounded sequence in \mathbb{R}^d and hence has a convergent subsequence, say $\{z_{n_k}\}_{k \geq 1}$ converging to z . Since C is closed and for every $k \geq 1$, $z_{n_k} \in C$, we have $z \in C$. Since the graph of H_x is closed, we have, $z \in H_x(s)$. Hence $H_x(s) \cap C \neq \emptyset$. Therefore $\{s \in \mathcal{S} : H_x(s) \cap C \neq \emptyset\}$ is closed and clearly belongs to $\mathcal{B}(\mathcal{S})$, which establishes the measurability of H_x .

- (ii) Fix some $l \geq 1$. For every $s \in \mathcal{S}$, the convexity and compactness of $H_x^{(l)}(s)$ follows from the definition of $H_x^{(l)}$ and Lemma 3.2(i). Boundedness of $H_x^{(l)}$ follows from the definition of $H_x^{(l)}$ and Lemma 3.2(iii). Let $\{s_n\}_{n \geq 1}$ be a sequence in \mathcal{S} converging to s and $\{z_n \in H_x^{(l)}(s_n)\}_{n \geq 1}$ converging to z . Using the continuity of $H^{(l)}$ and the definition of $H_x^{(l)}$, we have, for every $\epsilon > 0$, there exists $\delta > 0$ such that $\bigcup_{\{n: d(s_n, s) < \delta\}} H_x^{(l)}(s_n) \subseteq H_x^{(l)}(s) + \epsilon U$. Since $H_x^{(l)}(s) + \epsilon U$ is closed, we have, $\bigcup_{\{n: d(s_n, s) < \delta\}} H_x^{(l)}(s_n) \subseteq H_x^{(l)}(s) + \epsilon U$. Hence for every $\epsilon > 0$, $z \in H_x^{(l)}(s) + \epsilon U$. Since $H_x^{(l)}(s)$ is compact we have, $z \in H_x^{(l)}(s)$. Hence the graph of $H_x^{(l)}$ is closed and by argument similar to the one in part (i) of this proof we can prove the measurability of $H_x^{(l)}$.
- (iii) Since H_x is bounded, for every $f \in \mathcal{S}(H_x)$, we have $\sup_{s \in \mathcal{S}} \|f(s)\| \leq C_x$. Hence $\int_{\mathcal{S}} \|f(s)\| \mu(ds) \leq C_x$.
- (iv) Argument is same as in part (iii) of this proof.
- (v) Follows from definition of $h_x^{(l)}$ and Lemma 3.3. ■

Recall that $\mathcal{P}(\mathcal{S} \times U)$ denotes the set of probability measures on $(\mathcal{S} \times U, \mathcal{B}(\mathcal{S}) \otimes \mathcal{B}(U))$ with the Prohorov's topology. For any probability measure $\nu \in \mathcal{P}(\mathcal{S} \times U)$, let $\nu_{\mathcal{S}} \in \mathcal{P}(\mathcal{S})$ denote the image of ν under the projection, $\mathcal{S} \times U \rightarrow \mathcal{S}$ (i.e. for every $A \in \mathcal{B}(\mathcal{S})$, $\nu_{\mathcal{S}}(A) := \int_{A \times U} \nu(ds, du)$). The next lemma provides a useful characterization of the integral of set valued maps $H_x^{(l)}$ in terms of their parametrization, $h_x^{(l)}$.

Lemma 4.2 *For every $x \in \mathbb{R}^d$, for every $l \geq 1$, for every probability measure μ on $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$,*

$$\int_{\mathcal{S}} H_x^{(l)}(s) \mu(ds) = \left\{ \int_{\mathcal{S} \times U} h_x^{(l)}(s, u) \nu(ds, du) : \nu \in \mathcal{P}(\mathcal{S} \times U), \nu_{\mathcal{S}} = \mu \right\}$$

Proof Fix some $x \in \mathbb{R}^d$, $l \geq 1$ and $\mu \in \mathcal{P}(\mathcal{S})$.

Let $z \in \int_{\mathcal{S}} H_x^{(l)}(s) \mu(ds)$. There exists $f \in \mathcal{S}(H_x^{(l)})$ such that, $z = \int_{\mathcal{S}} f(s) \mu(ds)$. Define a set valued map, $G : \mathcal{S} \rightarrow \{\text{subsets of } U\}$ such that, for every $s \in \mathcal{S}$, $G(s) := \{u \in U : h^{(l)}(x, s, u) = f(s)\}$. Clearly for every $s \in \mathcal{S}$, $G(s)$ is non-empty and from the continuity of $h^{(l)}$ we have that $G(s)$ is a closed subset of U . Moreover G is a measurable set valued map (the proof of measurability has been provided in appendix B). Since G is measurable, by Lemma 3.5(i) we have, $\mathcal{S}(G) \neq \emptyset$. For any $g \in \mathcal{S}(G)$, we have, for every $s \in \mathcal{S}$, $h^{(l)}(x, s, g(s)) = f(s)$. Define $\hat{g} : \mathcal{S} \rightarrow \mathcal{S} \times U$ such that, for every $s \in \mathcal{S}$, $\hat{g}(s) = (s, g(s))$. Let $\nu := \mu \hat{g}^{-1}$ (push-forward measure). Clearly $\nu_{\mathcal{S}} = \mu$ and $\int_{\mathcal{S} \times U} h^{(l)}(x, s, u) \nu(ds, du) = \int_{\mathcal{S} \times U} h^{(l)}(x, s, u) \mu \hat{g}^{-1}(ds, du) = \int_{\mathcal{S}} h^{(l)}(x, \hat{g}(s)) \mu(ds) = \int_{\mathcal{S}} h^{(l)}(x, s, g(s)) \mu(ds) = \int_{\mathcal{S}} f(s) \mu(ds) = z$.

Let $\nu \in \mathcal{P}(\mathcal{S} \times U)$ such that, $\nu_{\mathcal{S}} = \mu$. By Corollary 3.1.2 in [8], there exists a μ -a.s. unique measurable map $q : \mathcal{S} \rightarrow \mathcal{P}(U)$ such that, $\nu(ds, du) = \mu(ds) q(s, du)$. By definition of $h_x^{(l)}$ and Lemma 3.3, we have, for every $s \in \mathcal{S}$, $H_x^{(l)}(s) = h_x^{(l)}(s, U) = \left\{ h_x^{(l)}(s, u) : u \in U \right\}$ is convex and compact subset of \mathbb{R}^d . Hence for every $s \in \mathcal{S}$, $\int_U h_x^{(l)}(s, u) q(s, du) \in H_x^{(l)}(s)$. Let $f : \mathcal{S} \rightarrow \mathbb{R}^d$ be such that for every $s \in \mathcal{S}$, $f(s) = \int_U h_x^{(l)}(s, u) q(s, du)$. Then, clearly f is measurable and $f \in \mathcal{S}(H_x^{(l)})$. Therefore, $\int_{\mathcal{S} \times U} h_x^{(l)}(s, u) \nu(ds, du) = \int_{\mathcal{S}} \left[\int_U h_x^{(l)}(s, u) q(s, du) \right] \mu(ds) = \int_{\mathcal{S}} f(s) \mu(ds) \in \int_{\mathcal{S}} H_x^{(l)}(s) \mu(ds)$. ■

Recall that for every $x \in \mathbb{R}^d$, $D(x)$ denotes the set of stationary distributions associated with the Markov chain whose transition probability kernel is given by $\Pi(x, \cdot)(\cdot)$ as in assumption (A2). We shall now define the set valued maps associated with the limiting differential inclusion which is obtained by averaging the set valued map H w.r.t. the stationary distributions of the Markov noise. We shall perform a similar operation on set valued maps $H^{(l)}$ which approximate H .

Definition :

(i) Define $\hat{H} : \mathbb{R}^d \rightarrow \{\text{subsets of } \mathbb{R}^d\}$ such that, for every $x \in \mathbb{R}^d$,

$$\hat{H}(x) := \cup_{\mu \in D(x)} \int_S H_x(s) \mu(ds). \quad (5)$$

(ii) For every $l \geq 1$, define $\hat{H}^{(l)} : \mathbb{R}^d \rightarrow \{\text{subsets of } \mathbb{R}^d\}$ such that, for every $x \in \mathbb{R}^d$,

$$\hat{H}^{(l)}(x) := \cup_{\mu \in D(x)} \int_S H_x^{(l)}(s) \mu(ds). \quad (6)$$

For every $l \geq 1$, the differential inclusion (DI) associated with the set valued map $\hat{H}^{(l)}$ is given by,

$$\frac{dx}{dt} \in \hat{H}^{(l)}(x). \quad (7)$$

Similarly the differential inclusion (DI) associated with the set valued map \hat{H} is given by,

$$\frac{dx}{dt} \in \hat{H}(x). \quad (8)$$

A function $\mathbf{x} : \mathbb{R} \rightarrow \mathbb{R}^d$ is said to be a **solution** of DI (7) (or DI (8)) with initial condition $x_0 \in \mathbb{R}^d$ if, \mathbf{x} is absolutely continuous, $\mathbf{x}(0) = x_0$ and for a.e. $t \in \mathbb{R}$, $\frac{d\mathbf{x}(t)}{dt} \in \hat{H}^{(l)}(\mathbf{x}(t))$ (or $\frac{d\mathbf{x}(t)}{dt} \in \hat{H}(\mathbf{x}(t))$).

For every $l \geq 1$, for every initial condition $x_0 \in \mathbb{R}^d$, let $\Sigma^{(l)}(x_0)$ denote the set of solutions to DI (7) with initial condition x_0 , that is,

$$\Sigma^{(l)}(x_0) = \left\{ \mathbf{x} : \mathbb{R} \rightarrow \mathbb{R}^d : \mathbf{x} \text{ is a solution of DI (7), } \mathbf{x}(0) = x_0 \right\}.$$

Similarly, let $\Sigma(x_0)$ denote the set of solutions of DI (8) with initial condition $x_0 \in \mathbb{R}^d$. For any $A \subseteq \mathbb{R}^d$, for every $l \geq 1$, $\Sigma^{(l)}(A) := \cup_{x_0 \in A} \Sigma^{(l)}(x_0)$ (similarly, $\Sigma(A) := \cup_{x_0 \in A} \Sigma(x_0)$).

The next lemma establishes some properties of the set valued map $\hat{H}^{(l)}$ which ensure the existence of solutions to DI (7).

Lemma 4.3 *For every $l \geq 1$, $\hat{H}^{(l)}$ satisfies the following,*

(i) *for $K^{(l)}$ as in Lemma 3.2(iii), we have, for every $x \in \mathbb{R}^d$,*

$$\sup_{z \in \hat{H}^{(l)}(x)} \|z\| \leq K^{(l)}(1 + \|x\|).$$

(ii) *for every $x \in \mathbb{R}^d$, $\hat{H}^{(l)}(x)$ is a non-empty, convex and compact subset of \mathbb{R}^d .*

(iii) *for every sequence $\{x_n\}_{n \geq 1}$ in \mathbb{R}^d converging to x , for every sequence $\{z_n \in \hat{H}^{(l)}(x_n)\}_{n \geq 1}$ converging to z , we have, $z \in \hat{H}^{(l)}(x)$.*

Proof Fix $l \geq 1$.

(i) For every $x \in \mathbb{R}^d$, from Lemma 4.1(ii), we have, for every $f \in \mathcal{S}(H_x^{(l)})$, for every $\mu \in D(x)$, $\|\int_S f(s) \mu(ds)\| \leq \int_S \|f(s)\| \mu(ds) \leq C_x^{(l)} = K^{(l)}(1 + \|x\|)$. Using the above and from the definition of $\hat{H}^{(l)}$ we have, $\sup_{z \in \hat{H}^{(l)}(x)} \|z\| = \sup_{\mu \in D(x)} \sup_{f \in \mathcal{S}(H_x^{(l)})} \|\int_S f(s) \mu(ds)\| \leq K^{(l)}(1 + \|x\|)$.

- (ii) For every $x \in \mathbb{R}^d$, $\hat{H}^{(l)}(x)$ is clearly non-empty. Let $z_1, z_2 \in \hat{H}^{(l)}(x)$ and $\theta \in (0, 1)$. By definition of $\hat{H}^{(l)}$ and by Lemma 4.2, we have that there exist $\nu^1, \nu^2 \in \mathcal{P}(\mathcal{S} \times U)$ such that, for $i \in \{1, 2\}$, $z_i = \int_{\mathcal{S} \times U} h_x^{(l)}(s, u) \nu^i(ds, du)$ and $\nu_S^i \in D(x)$. Hence,

$$\begin{aligned} \theta z_1 + (1 - \theta) z_2 &= \theta \int_{\mathcal{S} \times U} h_x^{(l)}(s, u) \nu^1(ds, du) + (1 - \theta) \int_{\mathcal{S} \times U} h_x^{(l)}(s, u) \nu^2(ds, du) \\ &= \int_{\mathcal{S} \times U} h_x^{(l)}(s, u) (\theta \nu^1 + (1 - \theta) \nu^2)(ds, du), \end{aligned}$$

where, $\theta \nu^1 + (1 - \theta) \nu^2 \in \mathcal{P}(\mathcal{S} \times U)$ such that, for every $A \in \mathcal{B}(\mathcal{S}) \otimes \mathcal{B}(U)$, $(\theta \nu^1 + (1 - \theta) \nu^2)(A) = \theta \nu^1(A) + (1 - \theta) \nu^2(A)$. Further $(\theta \nu^1 + (1 - \theta) \nu^2)_S = \theta \nu_S^1 + (1 - \theta) \nu_S^2$ and it belongs to $D(x)$, since for every $x \in \mathbb{R}^d$, $D(x)$ is convex.

By part (i) of this proof, for every $x \in \mathbb{R}^d$, $\hat{H}^{(l)}(x)$ is bounded. Hence in order to show compactness it is enough to prove that for every $x \in \mathbb{R}^d$, $\hat{H}^{(l)}(x)$ is closed. Let $\{z_n\}_{n \geq 1}$ be a sequence in $\hat{H}^{(l)}(x)$ for some $x \in \mathbb{R}^d$, converging to z . By Lemma 4.2, we know that for every $n \geq 1$, there exists $\nu^n \in \mathcal{P}(\mathcal{S} \times U)$ such that, $z_n = \int_{\mathcal{S} \times U} h_x^{(l)}(s, u) \nu^n(ds, du)$ and $\nu_S^n \in D(x)$. Since $\mathcal{S} \times U$ is a compact metric space, by Prohorov's theorem (see theorem 2.3.1 in [8]), there exists a subsequence $\{n_k\}_{k \geq 1}$ such that, $\{\nu^{n_k}\}_{k \geq 1}$ converges to some $\nu \in \mathcal{P}(\mathcal{S} \times U)$. Clearly $\{z_{n_k}\}_{k \geq 1}$ converges to z and $\{\nu_S^{n_k}\}_{k \geq 1}$ converges to ν_S in $\mathcal{P}(\mathcal{S})$. Since for every $x \in \mathbb{R}^d$, $D(x)$ is a closed subset of $\mathcal{P}(\mathcal{S})$, we have $\nu_S \in D(x)$. By theorem 2.1.1(ii) in [8] and continuity of $h_x^{(l)}$ we have $\left\{ \int_{\mathcal{S} \times U} h_x^{(l)}(s, u) \nu^{n_k}(ds, du) \right\}_{k \geq 1}$ converges to $\int_{\mathcal{S} \times U} h_x^{(l)}(s, u) \nu(ds, du)$ and hence $z = \int_{\mathcal{S} \times U} h_x^{(l)}(s, u) \nu(ds, du)$ with $\nu_S \in D(x)$. Therefore $z \in \hat{H}^{(l)}(x)$ and hence $\hat{H}^{(l)}(x)$ is closed.

- (iii) Let $\{x_n\}_{n \geq 1}$ be a sequence in \mathbb{R}^d converging to x and $\{z_n \in \hat{H}^{(l)}(x_n)\}_{n \geq 1}$ converging to z . By Lemma 4.2, we have, for every $n \geq 1$, there exists $\nu^n \in \mathcal{P}(\mathcal{S} \times U)$, such that $z_n = \int_{\mathcal{S} \times U} h_{x_n}^{(l)}(s, u) \nu^n(ds, du)$ and $\nu_S^n \in D(x_n)$. Since $\mathcal{S} \times U$ is a compact metric space, by Prohorov's theorem we have that, there exists a subsequence, $\{n_k\}_{k \geq 1}$ such that, $\{\nu^{n_k}\}_{k \geq 1}$ converges to some $\nu \in \mathcal{P}(\mathcal{S} \times U)$. Since $\{x_{n_k}\}_{k \geq 1}$ converges to x , we have $\{h_{x_{n_k}}^{(l)}\}_{k \geq 1}$ converges uniformly to $h_x^{(l)}$ (for a proof see appendix C). By uniform convergence of $h_{x_{n_k}}^{(l)}$ to $h_x^{(l)}$ and Theorem 2.1.1(ii) in [8] we have that the sequence $\left\{ \int_{\mathcal{S} \times U} h_{x_{n_k}}^{(l)}(s, u) \nu^{n_k}(ds, du) \right\}_{k \geq 1}$ converges to $\int_{\mathcal{S} \times U} h_x^{(l)}(s, u) \nu(ds, du)$. Hence $z = \int_{\mathcal{S} \times U} h_x^{(l)}(s, u) \nu(ds, du)$. Clearly $\{\nu_S^{n_k}\}_{k \geq 1}$ converges to ν_S and by closed graph property of the map $x \rightarrow D(x)$, we have $\nu_S \in D(x)$. Hence $z = \int_{\mathcal{S} \times U} h_x^{(l)}(s, u) \nu(ds, du)$ and $\nu_S \in D(x)$. Therefore by Lemma 4.2 and definition of $\hat{H}^{(l)}$, we have, $z \in \hat{H}^{(l)}(x)$. \blacksquare

The set valued map satisfying properties stated in Lemma 4.3, is called a Marchaud map (see page 62 of [16]). By Lemma 4.3, we know that for every $l \geq 1$, the set valued maps $\hat{H}^{(l)}$ are Marchaud maps. For such maps their associated DIs are known to admit at least one solution through every initial condition (see [14], chapter 2.1 or [1], section 1.2). Thus for every $l \geq 1$, for every initial condition $x_0 \in \mathbb{R}^d$, $\Sigma^{(l)}(x_0) \neq \emptyset$.

The next lemma establishes relation between \hat{H} and $\hat{H}^{(l)}$.

Lemma 4.4 For every $x \in \mathbb{R}^d$,

(i) for every $l \geq 1$, $\hat{H}(x) \subseteq \hat{H}^{(l+1)}(x) \subseteq \hat{H}^{(l)}(x)$.

(ii) $\cup_{\mu \in D(x)} \cap_{l \geq 1} \int_{\mathcal{S}} H_x^{(l)}(s) \mu(ds) = \cap_{l \geq 1} \hat{H}^{(l)}(x)$.

$$(iii) \hat{H}(x) = \cap_{l \geq 1} \hat{H}^{(l)}(x).$$

Proof Fix $x \in \mathbb{R}^d$.

(i) Fix $l \geq 1$. By Lemma 3.2(ii), for every $(x, s) \in \mathbb{R}^d \times \mathcal{S}$, $H(x, s) \subseteq H^{(l+1)}(x, s) \subseteq H^{(l)}(x, s)$. Hence, for every $x \in \mathbb{R}^d$, by definition of H_x and $H_x^{(l)}$ we have, for every $s \in \mathcal{S}$, $H_x(s) \subseteq H_x^{(l+1)}(s) \subseteq H_x^{(l)}(s)$. Therefore, $\mathcal{S}(H_x) \subseteq \mathcal{S}(H_x^{(l+1)}) \subseteq \mathcal{S}(H_x^{(l)})$. From the above we obtain that, for every $\mu \in D(x)$, $\int_{\mathcal{S}} H_x(s) \mu(ds) \subseteq \int_{\mathcal{S}} H_x^{(l+1)}(s) \mu(ds) \subseteq \int_{\mathcal{S}} H_x^{(l)}(s) \mu(ds)$ which leads to the desired conclusion.

(ii) Clearly by definition of the set valued map $\hat{H}^{(l)}$ we have that $\cup_{\mu \in D(x)} \cap_{l \geq 1} \int_{\mathcal{S}} H_x^{(l)}(s) \mu(ds) \subseteq \cap_{l \geq 1} \hat{H}^{(l)}(x)$. Let $z \in \cap_{l \geq 1} \hat{H}^{(l)}(x)$. Then for every $l \geq 1$, there exists $\mu^{(l)} \in D(x)$ such that $z \in \int_{\mathcal{S}} H_x^{(l)}(s) \mu^{(l)}(ds)$. By Prohorov's theorem the sequence of probability measures $\{\mu^{(l)}\}_{l \geq 1}$ is relatively compact in $\mathcal{P}(\mathcal{S})$ and hence has a limit point, say μ . Let $\{\mu^{(l_k)}\}_{k \geq 1}$ denote a subsequence with μ as its limit and since $D(x)$ is a closed subset of $\mathcal{P}(\mathcal{S})$, we have that $\mu \in D(x)$. For every $l \geq 1$, for every k such that $l_k \geq l$, we have that $\mathcal{S}(H_x^{(l_k)}) \subseteq \mathcal{S}(H_x^{(l)})$. Thus for every $l \geq 1$, for every k such that $l_k \geq l$, we have that $z \in \int_{\mathcal{S}} H_x^{(l)}(s) \mu^{(l_k)}(ds)$. By Lemma 4.2, we have that for every $l \geq 1$, for every k such that $l_k \geq l$, there exists $\nu^{(l, l_k)} \in \mathcal{P}(\mathcal{S} \times U)$ with $\nu_{\mathcal{S}}^{(l, l_k)} = \mu^{(l_k)}$ and $z = \int_{\mathcal{S} \times U} h^{(l)}(x, s, u) \nu^{(l, l_k)}(ds, du)$.

Fix $l \geq 1$. Since $\mathcal{S} \times U$ is a compact metric space, by Prohorov's theorem we have that the sequence $\{\nu^{(l, l_k)}\}_{k \geq 1}$ is relatively compact in $\mathcal{P}(\mathcal{S} \times U)$. Let $\nu^{(l)}$ denote a limit point of the above sequence and its clear that $\nu_{\mathcal{S}}^{(l)} = \mu$. Then clearly $z = \int_{\mathcal{S} \times U} h^{(l)}(x, s, u) \nu^{(l)}(ds, du)$ and hence by Lemma 4.2, $z \in \int_{\mathcal{S}} H_x^{(l)}(s) \mu(ds)$. Since $l \geq 1$ is arbitrary, we get that for $\mu \in D(x)$, for every $l \geq 1$, $z \in \int_{\mathcal{S}} H_x^{(l)}(s) \mu(ds)$. Thus $z \in \cup_{\mu \in D(x)} \cap_{l \geq 1} \int_{\mathcal{S}} H_x^{(l)}(s) \mu(ds)$ and hence $\cap_{l \geq 1} \hat{H}^{(l)}(x) \subseteq \cup_{\mu \in D(x)} \cap_{l \geq 1} \int_{\mathcal{S}} H_x^{(l)}(s) \mu(ds)$.

(iii) By part(i) of this lemma we have, for every $x \in \mathbb{R}^d$, $\hat{H}(x) \subseteq \cap_{l \geq 1} \hat{H}^{(l)}(x)$. Therefore its enough to show that for every $x \in \mathbb{R}^d$, $\cap_{l \geq 1} \hat{H}^{(l)}(x) \subseteq \hat{H}(x)$. Fix $x \in \mathbb{R}^d$ and $\mu \in D(x)$. Let $z \in \cap_{l \geq 1} \int_{\mathcal{S}} H_x^{(l)}(s) \mu(ds)$. For every $l \geq 1$, there exists $f^{(l)} \in \mathcal{S}(H_x^{(l)})$ such that $z = \int_{\mathcal{S}} f^{(l)}(s) \mu(ds)$. Let $d(z, \int_{\mathcal{S}} H_x(s) \mu(ds)) := \inf \{ \|z - y\| : y \in \int_{\mathcal{S}} H_x(s) \mu(ds) \}$. Then, for every $l \geq 1$,

$$\begin{aligned} d(z, \int_{\mathcal{S}} H_x(s) \mu(ds)) &= \inf_{f \in \mathcal{S}(H_x)} \left\| \int_{\mathcal{S}} (f^{(l)}(s) - f(s)) \mu(ds) \right\| \\ &\leq \inf_{f \in \mathcal{S}(H_x)} \int_{\mathcal{S}} \|f^{(l)}(s) - f(s)\| \mu(ds) \\ &= \int_{\mathcal{S}} \inf_{f \in \mathcal{S}(H_x)} \|f^{(l)}(s) - f(s)\| \mu(ds), \end{aligned}$$

where the last equality follows from Lemma 1.3.12 in [15]. By Lemma 4.1(i), H_x is a measurable set valued map and hence by Lemma 3.5(ii), H_x admits a Castaing representation, $\{f_n\}_{n \geq 1}$. Hence for every $s \in \mathcal{S}$, $d(f^{(l)}(s), H_x(s)) \leq \inf_{f \in \mathcal{S}(H_x)} \|f^{(l)}(s) - f(s)\| \leq \inf_{n \geq 1} \|f^{(l)}(s) - f_n(s)\| = d(f^{(l)}(s), H_x(s))$ where $s \rightarrow d(f^{(l)}(s), H_x(s))$ is as in Lemma 3.7. Therefore, for every $l \geq 1$,

$$d(z, \int_{\mathcal{S}} H_x(s) \mu(ds)) \leq \int_{\mathcal{S}} d(f^{(l)}(s), H_x(s)) \mu(ds).$$

By observation (b) stated after Lemma 3.2, we have that, for every $s \in \mathcal{S}$, $\lim_{l \rightarrow \infty} d(f^{(l)}(s), H_x(s)) = 0$. Further by observation (a) stated after Lemma 3.2, Lemma 4.1(i) – (ii), we have that, for every $s \in \mathcal{S}$, $d(f^{(l)}(s), H_x(s)) \leq (\tilde{K} + K)(1 + \|x\|)$. Thus,

by bounded convergence theorem we get, $\lim_{l \rightarrow \infty} \int_{\mathcal{S}} d(f^{(l)}(s), H_x(s)) \mu(ds) = 0$. Hence, $d(z, \int_{\mathcal{S}} H_x(s) \mu(ds)) = 0$. By Lemma 3.6, we have that $\int_{\mathcal{S}} H_x(s) \mu(ds)$ is closed. Therefore $z \in \int_{\mathcal{S}} H_x(s) \mu(ds)$.

From the arguments in the previous paragraph, we have that for every $x \in \mathbb{R}^d$, for every $\mu \in D(x)$, $\cap_{l \geq 1} \int_{\mathcal{S}} H_x^{(l)}(s) \mu(ds) \subseteq \int_{\mathcal{S}} H_x(s) \mu(ds)$. Thus, for every $x \in \mathbb{R}^d$, $\cup_{\mu \in D(x)} \cap_{l \geq 1} \int_{\mathcal{S}} H_x^{(l)}(s) \mu(ds) \subseteq \cup_{\mu \in D(x)} \int_{\mathcal{S}} H_x(s) \mu(ds)$. Therefore, by part (ii) of this lemma we have that for every $x \in \mathbb{R}^d$, $\cap_{l \geq 1} \hat{H}^{(l)}(x) \subseteq \hat{H}(x)$. ■

In the next lemma we show that the set valued map \hat{H} is a Marchaud map.

Lemma 4.5 *The set valued map \hat{H} satisfies the following:*

- (i) For $K > 0$ as in Lemma 4.1(i), for every $x \in \mathbb{R}^d$, $\sup_{z \in \hat{H}(x)} \|z\| \leq K(1 + \|x\|)$.
- (ii) For every $x \in \mathbb{R}^d$, $\hat{H}(x)$ is a non-empty, convex and compact subset of \mathbb{R}^d .
- (iii) For every sequence $\{x_n\}_{n \geq 1}$ in \mathbb{R}^d converging to x , for every sequence $\{z_n \in \hat{H}(x_n)\}_{n \geq 1}$ converging to z , we have, $z \in \hat{H}(x)$.

Proof :

- (i) For every $x \in \mathbb{R}^d$, from Lemma 4.1(i), we have, for every $f \in \mathcal{S}(H_x)$, for every $\mu \in D(x)$, $\|\int_{\mathcal{S}} f(s) \mu(ds)\| \leq \int_{\mathcal{S}} \|f(s)\| \mu(ds) \leq C_x = K(1 + \|x\|)$. Using the above and from the definition of \hat{H} we have, $\sup_{z \in \hat{H}(x)} \|z\| = \sup_{\mu \in D(x)} \sup_{f \in \mathcal{S}(H_x)} \|\int_{\mathcal{S}} f(s) \mu(ds)\| \leq K(1 + \|x\|)$.
- (ii) For every $x \in \mathbb{R}^d$, by Lemma 4.1(iii), for every $\mu \in D(x)$, H_x is μ -integrable. Hence \hat{H}_x is non-empty. By Lemma 4.3(ii) and Lemma 4.4(iii), we have that, $\hat{H}(x)$ is a convex and compact subset of \mathbb{R}^d .
- (iii) Let $\{x_n\}_{n \geq 1}$ be a sequence in \mathbb{R}^d , converging to x and $\{z_n \in \hat{H}(x_n)\}_{n \geq 1}$ converging to z . By Lemma 4.4(i), for every $l \geq 1$, for every $n \geq 1$, $z_n \in \hat{H}^{(l)}(x_n)$. Hence by Lemma 4.3(iii), we have that, for every $l \geq 1$, $z \in \hat{H}^{(l)}(x)$. Therefore by Lemma 4.4(iii), we have, $z \in \hat{H}(x)$. ■

By Lemma 4.5, we know that the set valued map \hat{H} is a Marchaud map. Hence for every $x_0 \in \mathbb{R}^d$, $\Sigma(x_0) \neq \emptyset$.

By definition, $\Sigma^{(l)}(x_0)$ and $\Sigma(x_0)$ are subsets of $\mathcal{C}(\mathbb{R}, \mathbb{R}^d)$, the space of all \mathbb{R}^d valued continuous functions on \mathbb{R} . $\mathcal{C}(\mathbb{R}, \mathbb{R}^d)$ is a complete metric space for distance \mathbf{D} , defined by,

$$\mathbf{D}(\mathbf{x}, \mathbf{z}) := \sum_{k=1}^{\infty} \frac{1}{2^k} \min(\|\mathbf{x} - \mathbf{z}\|_{[-k, k]}, 1),$$

where $\|\cdot\|_{[-k, k]}$ denotes the sup norm on $\mathcal{C}([-k, k], \mathbb{R}^d)$.

The next lemma summarizes some properties of the solutions of DI (7),(8) and the relation between them.

Lemma 4.6

- (i) For every $A \subseteq \mathbb{R}^d$ compact, $\Sigma(A)$ and for every $l \geq 1$, $\Sigma^{(l)}(A)$ are compact subsets of $\mathcal{C}(\mathbb{R}, \mathbb{R}^d)$.
- (ii) For every $x_0 \in \mathbb{R}^d$, for every $l \geq 1$, $\Sigma(x_0) \subseteq \Sigma^{(l+1)}(x_0) \subseteq \Sigma^{(l)}(x_0)$.

(iii) For every $x_0 \in \mathbb{R}^d$, $\Sigma(x_0) = \cap_{l \geq 1} \Sigma^{(l)}(x_0)$.

Proof :

(i) Follows from Lemma 3.1 in [1].

(ii) Fix some $x_0 \in \mathbb{R}^d$ and $l \geq 1$. Let $\mathbf{x} \in \Sigma(x_0)$. Then for a.e. $t \in \mathbb{R}$, $\frac{d\mathbf{x}(t)}{dt} \in \hat{H}(\mathbf{x}(t))$. By Lemma 4.4(i), we have, for every $t \geq 0$, $\hat{H}(\mathbf{x}(t)) \subseteq \hat{H}^{(l)}(\mathbf{x}(t))$. Therefore, for a.e. $t \in \mathbb{R}$, $\frac{d\mathbf{x}(t)}{dt} \in \hat{H}^{(l)}(\mathbf{x}(t))$. Hence, $\mathbf{x} \in \Sigma^{(l)}(x_0)$. A similar argument gives us that, $\Sigma^{(l+1)}(x_0) \subseteq \Sigma^{(l)}(x_0)$.

(iii) By part (ii) of this lemma we have that, for every $x_0 \in \mathbb{R}^d$, $\Sigma(x_0) \subseteq \cap_{l \geq 1} \Sigma^{(l)}(x_0)$. Therefore, it is enough to show that, for every $x_0 \in \mathbb{R}^d$, $\cap_{l \geq 1} \Sigma^{(l)}(x_0) \subseteq \Sigma(x_0)$. Let $\mathbf{x} \in \cap_{l \geq 1} \Sigma^{(l)}(x_0)$ for some $x_0 \in \mathbb{R}^d$. Then for every $l \geq 1$, for a.e. $t \in \mathbb{R}$, $\frac{d\mathbf{x}(t)}{dt} \in \hat{H}^{(l)}(\mathbf{x}(t))$. Thus for a.e. $t \in \mathbb{R}$, $\frac{d\mathbf{x}(t)}{dt} \in \cap_{l \geq 1} \hat{H}^{(l)}(\mathbf{x}(t))$. By Lemma 4.4(iii), we have that, for a.e. $t \in \mathbb{R}$, $\frac{d\mathbf{x}(t)}{dt} \in \hat{H}(\mathbf{x}(t))$. Hence $\mathbf{x} \in \Sigma(x_0)$. ■

5 Space of probability measure valued functions

In this section we define the space of probability measure valued measurable functions on $[0, \infty)$ and introduce an appropriate topology on this space. Such spaces are used in the theory of existence of optimal control for diffusions and can also be found in [7] and [17]. We shall use the following in the analysis of recursion (4) presented in the next section.

Let $\mathcal{M}(\mathcal{S} \times U)$ denote the set of all functions $\gamma : [0, \infty) \rightarrow \mathcal{P}(\mathcal{S} \times U)$, measurable. Formally,

$$\mathcal{M}(\mathcal{S} \times U) := \{\gamma : [0, \infty) \rightarrow \mathcal{P}(\mathcal{S} \times U) : \gamma \text{ is measurable}\}.$$

Similarly, define $\mathcal{M}(\mathcal{S})$, the set of all functions $\gamma : [0, \infty) \rightarrow \mathcal{P}(\mathcal{S})$, measurable. Formally,

$$\mathcal{M}(\mathcal{S}) := \{\gamma : [0, \infty) \rightarrow \mathcal{P}(\mathcal{S}) : \gamma \text{ is measurable}\}.$$

Let $\tau_{\mathcal{S} \times U}$ denote the topology on $\mathcal{M}(\mathcal{S} \times U)$ which is the coarsest topology that renders continuous the maps $\gamma \in \mathcal{M}(\mathcal{S} \times U) \rightarrow \int_0^T g(t) [\int_{\mathcal{S} \times U} f(s, u) \gamma(t)(ds, du)] dt \in \mathbb{R}$ for all $f \in \mathcal{C}(\mathcal{S} \times U, \mathbb{R})$, for all $T > 0$ and for all $g \in \mathbb{L}_2[0, T]$.

Similarly let $\tau_{\mathcal{S}}$ denote the topology on $\mathcal{M}(\mathcal{S})$ which is the coarsest topology that renders continuous the maps $\gamma \in \mathcal{M}(\mathcal{S}) \rightarrow \int_0^T g(t) [\int_{\mathcal{S}} f(s) \gamma(t)(ds)] dt \in \mathbb{R}$ for all $f \in \mathcal{C}(\mathcal{S}, \mathbb{R})$, for all $T > 0$ and for all $g \in \mathbb{L}_2[0, T]$.

The next result is a well known metrization lemma for the above defined topological spaces.

Lemma 5.1 [Metrization]

(i) The topological space $(\mathcal{M}(\mathcal{S} \times U), \tau_{\mathcal{S} \times U})$ is compact metrizable.

(ii) The topological space $(\mathcal{M}(\mathcal{S}), \tau_{\mathcal{S}})$ is compact metrizable.

For a proof of the above lemma we refer the reader to Lemma 2.1 in [7]. The next lemma provides a continuous map from $\mathcal{M}(\mathcal{S} \times U)$ to $\mathcal{M}(\mathcal{S})$ which is used later in this paper. Recall that for any probability measure, $\nu \in \mathcal{P}(\mathcal{S} \times U)$, $\nu_{\mathcal{S}}$ denotes the image of ν under projection, $\mathcal{S} \times U \rightarrow \mathcal{S}$ (i.e. $\nu_{\mathcal{S}} \in \mathcal{P}(\mathcal{S})$ such that for every $A \in \mathcal{B}(\mathcal{S})$, $\nu_{\mathcal{S}}(A) = \int_{A \times U} \nu(ds, du)$).

Lemma 5.2 :

(i) The map $\lambda : \mathcal{P}(\mathcal{S} \times U) \rightarrow \mathcal{P}(\mathcal{S})$ such that, for every $\nu \in \mathcal{P}(\mathcal{S} \times U)$, $\lambda(\nu) := \nu_{\mathcal{S}}$, is continuous.

(ii) For every $\gamma \in \mathcal{M}(\mathcal{S} \times U)$, $\lambda \circ \gamma \in \mathcal{M}(\mathcal{S})$.

(iii) The map $\Lambda : \mathcal{M}(\mathcal{S} \times U) \rightarrow \mathcal{M}(\mathcal{S})$ such that, for every $\gamma \in \mathcal{M}(\mathcal{S} \times U)$, $\Lambda(\gamma) := \lambda \circ \gamma$, is continuous.

Proof :

- (i) Let $\{\nu^n\}_{n \geq 1}$ be a sequence in $\mathcal{P}(\mathcal{S} \times U)$ converging to $\nu \in \mathcal{P}(\mathcal{S} \times U)$. Let $\pi : \mathcal{S} \times U \rightarrow \mathcal{S}$ denote the projection map. Then for every $f \in \mathcal{C}(\mathcal{S}, \mathbb{R})$, $f \circ \pi \in \mathcal{C}(\mathcal{S} \times U, \mathbb{R})$. By Theorem 2.1.1 (ii) in [8] and the compactness of $\mathcal{S} \times U$, for every $f \in \mathcal{C}(\mathcal{S}, \mathbb{R})$, $\int_{\mathcal{S} \times U} (f \circ \pi)(s, u) \nu^n(ds, du) \rightarrow \int_{\mathcal{S} \times U} (f \circ \pi)(s, u) \nu(ds, du)$ as $n \rightarrow \infty$. Hence for every $f \in \mathcal{C}(\mathcal{S}, \mathbb{R})$, $\int_{\mathcal{S}} f(s) \nu^n \pi^{-1}(ds) \rightarrow \int_{\mathcal{S}} f(s) \nu \pi^{-1}(ds)$. Observing that for every $n \geq 1$, $\nu^n \pi^{-1} = \nu_{\mathcal{S}}^n$ and $\nu \pi^{-1} = \nu_{\mathcal{S}}$ gives us that, for every $f \in \mathcal{C}(\mathcal{S}, \mathbb{R})$, $\int_{\mathcal{S}} f(s) \nu_{\mathcal{S}}^n(ds) \rightarrow \int_{\mathcal{S}} f(s) \nu_{\mathcal{S}}(ds)$. Hence $\lambda(\nu^n) \rightarrow \lambda(\nu)$ in $\mathcal{P}(\mathcal{S})$, which gives us continuity of $\lambda(\cdot)$.
- (ii) Composition of two measurable functions is measurable.
- (iii) Let $\{\gamma_n\}_{n \geq 1}$ be a sequence in $\mathcal{M}(\mathcal{S} \times U)$ converging to $\gamma \in \mathcal{M}(\mathcal{S} \times U)$. Then, for every $f \in \mathcal{C}(\mathcal{S} \times U, \mathbb{R})$, for every $T > 0$ and for every $g \in \mathbb{L}_2[0, T]$, $\int_0^T g(t) [\int_{\mathcal{S} \times U} f(s, u) \gamma_n(t)(ds, du)] dt \rightarrow \int_0^T g(t) [\int_{\mathcal{S} \times U} f(s, u) \gamma(t)(ds, du)] dt$ as $n \rightarrow \infty$. Hence, for every $f \in \mathcal{C}(\mathcal{S}, \mathbb{R})$, for every $T > 0$ and for every $g \in \mathbb{L}_2[0, T]$, $\int_0^T g(t) [\int_{\mathcal{S} \times U} (f \circ \pi)(s, u) \gamma_n(t)(ds, du)] dt \rightarrow \int_0^T g(t) [\int_{\mathcal{S} \times U} (f \circ \pi)(s, u) \gamma(t)(ds, du)] dt$ as $n \rightarrow \infty$. By argument similar to part (i) of this lemma, we have, for every $f \in \mathcal{C}(\mathcal{S}, \mathbb{R})$, for every $T > 0$ and for every $g \in \mathbb{L}_2[0, T]$, $\int_0^T g(t) [\int_{\mathcal{S}} f(s) (\lambda \circ \gamma_n)(t)(ds)] dt \rightarrow \int_0^T g(t) [\int_{\mathcal{S}} f(s) (\lambda \circ \gamma)(t)(ds)] dt$ as $n \rightarrow \infty$. Therefore $\{(\lambda \circ \gamma_n)\}_{n \geq 1}$ converges to $(\lambda \circ \gamma)$ in $\mathcal{M}(\mathcal{S})$, which gives us continuity of $\Lambda(\cdot)$. ■

6 Recursion analysis

Before we present the analysis of recursion (4) we begin with some preliminaries in the next subsection. Later we present the main result of the paper followed by the limit set theorem which characterizes the limit set of the recursion in terms of the dynamics of the limiting DI. Throughout this section we shall assume that assumptions (A1) – (A5) are satisfied.

6.1 Preliminaries

Define $t(0) := 0$ and for every $n \geq 1$, $t(n) := \sum_{k=0}^{n-1} a(k)$. Define the stochastic process $\bar{X} : \Omega \times [0, \infty) \rightarrow \mathbb{R}^d$ such that, for every $(\omega, t) \in \Omega \times [0, \infty)$,

$$\bar{X}(\omega, t) := \left(\frac{t - t(n)}{t(n+1) - t(n)} \right) X_{n+1}(\omega) + \left(\frac{t(n+1) - t}{t(n+1) - t(n)} \right) X_n(\omega),$$

where n is such that $t \in [t(n), t(n+1))$.

Recall that recursion (4) is given by,

$$X_{n+1} - X_n - a(n)M_{n+1} \in H(X_n, S_n),$$

for every $n \geq 0$. By Lemma 3.2, we have that, for every $l \geq 1$, for every $n \geq 0$, $H(X_n, S_n) \subseteq H^{(l)}(X_n, S_n)$. Therefore, for every $l \geq 1$, recursion (4) can be written as,

$$X_{n+1} - X_n - a(n)M_{n+1} \in H^{(l)}(X_n, S_n), \quad (9)$$

for every $n \geq 0$. By Lemma 3.3, we know that for every $l \geq 1$, the set valued map $H^{(l)}$, admits a parametrization. The next lemma allows us write the recursion in terms of the parametrization of $H^{(l)}$.

Lemma 6.1 For every $l \geq 1$, for every $n \geq 0$, there exists a U -valued random variable on Ω , say $U_n^{(l)}$, such that,

$$X_{n+1} - X_n - a(n)M_{n+1} = h^{(l)}(X_n, S_n, U_n^{(l)}),$$

where $h^{(l)}$ is as in Lemma 3.3.

Proof Fix $l \geq 1$ and $n \geq 0$. Define a set valued map $G : \Omega \rightarrow \{\text{subsets of } U\}$, such that, for every $\omega \in \Omega$, $G(\omega) := \{u \in U : h^{(l)}(X_n(\omega), S_n(\omega), u) = X_{n+1}(\omega) - X_n(\omega) - a(n)M_{n+1}(\omega)\}$. By Lemma 3.3 and by (9), we have that for every $\omega \in \Omega$, $G(\omega) \neq \emptyset$. By continuity of $h^{(l)}(\cdot)$, we have that for every $\omega \in \Omega$, $G(\omega)$ is closed. For any $C \subseteq U$, closed, $G^{-1}(C) = \{\omega \in \Omega : d(X_{n+1}(\omega) - X_n(\omega) - a(n)M_{n+1}(\omega), h^{(l)}(X_n(\omega), S_n(\omega), C)) = 0\}$. Clearly the set valued map $\omega \rightarrow h^{(l)}(X_n(\omega), S_n(\omega), C)$ is measurable and hence by Lemma 3.7, the map $\omega \rightarrow d(X_{n+1}(\omega) - X_n(\omega) - a(n)M_{n+1}(\omega), h^{(l)}(X_n(\omega), S_n(\omega), C))$ is measurable. Thus $G^{-1}(C) \in \mathcal{F}$. Therefore by Lemma 3.5, $\mathcal{S}(G) \neq \emptyset$. Set $U_n^{(l)}$ to be any measurable selection of G . ■

For every $l \geq 1$, define $\Gamma^{(l)} : \Omega \times [0, \infty) \rightarrow \mathcal{P}(\mathcal{S} \times U)$, such that, for every $(\omega, t) \in \Omega \times [0, \infty)$,

$$\Gamma^{(l)}(\omega, t) = \delta_{S_n(\omega)} \otimes \delta_{U_n^{(l)}(\omega)}, \quad (10)$$

where n is such that $t \in [t(n), t(n+1))$, $\delta_{S_n(\omega)} \in \mathcal{P}(\mathcal{S})$ denotes the Dirac measure in $\mathcal{P}(\mathcal{S})$ (i.e. for any $A \in \mathcal{B}(\mathcal{S})$, $\delta_{S_n(\omega)}(A) = 1$ if $S_n(\omega) \in A$ or 0 otherwise) and $\delta_{U_n^{(l)}(\omega)} \in \mathcal{P}(U)$ denotes the Dirac measure in $\mathcal{P}(U)$.

For every $l \geq 1$, for every $\nu \in \mathcal{P}(\mathcal{S} \times U)$, define $\tilde{h}_\nu^{(l)} : \mathbb{R}^d \rightarrow \mathbb{R}^d$, such that for every $x \in \mathbb{R}^d$, $\tilde{h}_\nu^{(l)}(x) := \int_{\mathcal{S} \times U} h^{(l)}(x, s, u) \nu(ds, du)$ where $h^{(l)}(\cdot)$ are as in Lemma 3.3. The next lemma provides an equicontinuity result used later.

Lemma 6.2 For every $l \geq 1$, for every $r > 0$, $\{\tilde{h}_\nu^{(l)}(\cdot)|_{rU} : \nu \in \mathcal{P}(\mathcal{S} \times U)\}$ is an equicontinuous family where $\tilde{h}_\nu^{(l)}(\cdot)|_{rU}$ denotes the restriction of $\tilde{h}_\nu^{(l)}(\cdot)$ to the closed ball of radius r .

Proof Fix $l \geq 1$ and $r > 0$. Since $rU \times \mathcal{S} \times U$ is compact, by Heine-Cantor theorem, $h^{(l)}(\cdot)|_{rU \times \mathcal{S} \times U}$ is uniformly continuous. Fix $\epsilon > 0$. Then, there exists $\delta > 0$ (depending on l, r and ϵ), such that, for every $x, x' \in rU$, for every $s, s' \in \mathcal{S}$, for every $u, u' \in U$ satisfying, $\|x - x'\| < \delta$, $d_{\mathcal{S}}(s, s') < \delta$, $\|u - u'\| < \delta$, we have, $\|h^{(l)}(x, s, u) - h^{(l)}(x', s', u')\| < \epsilon$. Therefore for δ as above, for $x, x' \in rU$ satisfying $\|x - x'\| < \delta$, we have that for every $\nu \in \mathcal{P}(\mathcal{S} \times U)$, $\|\tilde{h}_\nu^{(l)}(x) - \tilde{h}_\nu^{(l)}(x')\| = \|\int_{\mathcal{S} \times U} (h^{(l)}(x, s, u) - h^{(l)}(x', s, u)) \nu(ds, du)\| \leq \int_{\mathcal{S} \times U} \|h^{(l)}(x, s, u) - h^{(l)}(x', s, u)\| \nu(ds, du) < \int_{\mathcal{S} \times U} \epsilon \nu(ds, du) = \epsilon$. ■

For every $l \geq 1$, define the stochastic process, $Y^{(l)} : \Omega \times [0, \infty) \rightarrow \mathbb{R}^d$, such that, for every $(\omega, t) \in \Omega \times [0, \infty)$, $Y^{(l)}(\omega, t) := h^{(l)}(X_n(\omega), S_n(\omega), U_n^{(l)}(\omega))$, where n is such that $t \in [t(n), t(n+1))$.

In what follows, in the next two subsections, most of the arguments are sample path wise. We omit ω from our notation and use lower case letters to denote the above defined quantities along a particular sample path, for example, $\bar{x}(t)$, x_n , s_n , m_n , $u_n^{(l)}$, $y^{(l)}(t)$ and $\gamma^{(l)}(t)$ are to be understood as $\bar{X}(\omega, t)$, $X_n(\omega)$, $S_n(\omega)$, $M_n(\omega)$, $U_n^{(l)}(\omega)$, $Y^{(l)}(\omega, t)$ and $\Gamma^{(l)}(\omega, t)$ respectively for some ω fixed.

6.2 Main result - Asymptotic pseudotrajectory

For every $\omega \in \Omega$, for every $l \geq 1$, for every $\tilde{t} \geq 0$, let $\tilde{x}^{(l)}(\cdot; \tilde{t})$ denote the solution of the o.d.e.,

$$\dot{\tilde{x}}^{(l)}(t; \tilde{t}) = y^{(l)}(t + \tilde{t}), \quad (11)$$

for every $t \geq 0$, with initial condition $\tilde{x}^{(l)}(0; \tilde{t}) = \bar{x}(\tilde{t})$.

First we shall get rid of the additive noise terms, M_n . Let

$$\Omega_{a,s} := \{\omega \in \Omega : (A4) \text{ and } (A5) \text{ hold}\},$$

i.e., the set of sample paths where the iterates are stable and the additive noise terms are eventually negligible. By assumption (A4) and (A5), we have $\mathbb{P}(\Omega_{a,s}) = 1$.

Lemma 6.3 *For every $l \geq 1$, almost surely for every ω , for every $T > 0$,*

$$\lim_{t \rightarrow \infty} \sup_{0 \leq q \leq T} \|\bar{x}(t+q) - \tilde{x}^{(l)}(q; t)\| = 0.$$

Proof Fix $l \geq 1$, $\omega \in \Omega_{a,s}$ and $T > 0$. We shall prove the claim along the sequence $\{t(n)\}_{n \geq 0}$ as defined in section 6.1. The general claim easily follows from this special case.

Fix $n \geq 0$. Let $\tau(n, T) := \min \{k > n : t(k) \geq t(n) + T\}$. Let $q \in [0, T]$. Then, there exists k such that $t(n) + q \in [t(k), t(k+1))$ and $n \leq k \leq \tau(n, T) - 1$. By definition of $\bar{x}(\cdot)$ and $\tilde{x}^{(l)}(\cdot; t(n))$, we have that, $\bar{x}(t(n) + q) = \alpha x_k + (1 - \alpha)x_{k+1}$ and $\tilde{x}^{(l)}(q; t(n)) = \alpha \tilde{x}^{(l)}(t(k) - t(n); t(n)) + (1 - \alpha)\tilde{x}^{(l)}(t(k+1) - t(n); t(n))$ where $\alpha = \frac{t(k+1) - t(n) - q}{t(k+1) - t(k)}$. Since $\tilde{x}^{(l)}(\cdot; t(n))$ is a solution of the o.d.e. (11), we have that, for every $k \geq n$, $\tilde{x}^{(l)}(t(k) - t(n); t(n)) = x_n + \sum_{j=n}^{k-1} a(j)h^{(l)}(x_j, s_j, u_j^{(l)})$ and by Lemma 6.2, we have that, for every $k \geq n$, $x_k = \bar{x}(t(k)) = x_n + \sum_{j=n}^{k-1} a(j)h^{(l)}(x_j, s_j, u_j^{(l)}) + \sum_{j=n}^{k-1} a(j)m_{j+1}$. Thus,

$$\begin{aligned} \|\bar{x}(t(n) + q) - \tilde{x}^{(l)}(q; t(n))\| &\leq \left\| \alpha \sum_{j=n}^{k-1} a(j)m_{j+1} + (1 - \alpha) \sum_{j=n}^k a(j)m_{j+1} \right\| \\ &\leq \alpha \left\| \sum_{j=n}^{k-1} a(j)m_{j+1} \right\| + (1 - \alpha) \left\| \sum_{j=n}^k a(j)m_{j+1} \right\| \\ &\leq \sup_{n \leq k \leq \tau(n, T)} \left\| \sum_{j=n}^k a(j)m_{j+1} \right\|. \end{aligned}$$

Since the r.h.s. of the above inequality is independent of $q \in [0, T]$, we have, $\sup_{0 \leq q \leq T} \|\bar{x}(t(n) + q) - \tilde{x}^{(l)}(q; t(n))\| \leq \sup_{n \leq k \leq \tau(n, T)} \left\| \sum_{j=n}^k a(j)m_{j+1} \right\|$. Therefore, $\lim_{n \rightarrow \infty} \sup_{0 \leq q \leq T} \|\bar{x}(t(n) + q) - \tilde{x}^{(l)}(q; t(n))\| \leq \lim_{n \rightarrow \infty} \sup_{n \leq k \leq \tau(n, T)} \left\| \sum_{j=n}^k a(j)m_{j+1} \right\|$. Now the claim follows from assumption (A4). ■

Lemma 6.4 *For every $l \geq 1$, for almost every ω , $\{\tilde{x}^{(l)}(\cdot; t)\}_{t \geq 0}$ is relatively compact in $\mathcal{C}([0, \infty), \mathbb{R}^d)$.*

Proof Fix $l \geq 1$, $\omega \in \Omega_{a,s}$. By assumption (A5), we know that there exists $r > 0$ such that $\sup_{n \geq 0} \|x_n\| \leq r$ and hence $\sup_{t \geq 0} \tilde{x}^{(l)}(0; t) = \sup_{t \geq 0} \bar{x}(t) \leq r$.

For any $t \geq 0$, let $[t] := \max \{n \geq 0 : t(n) \leq t\}$. For every $t \geq 0$ and $q_1, q_2 \in [0, \infty)$ (w.l.o.g. assume $q_1 < q_2$) we have,

$$\begin{aligned} \|\tilde{x}^{(l)}(q_1; t) - \tilde{x}^{(l)}(q_2; t)\| &= \left\| \int_{q_1}^{q_2} h^{(l)}(x_{[t+q]}, s_{[t+q]}, u_{[t+q]}^{(l)}) dq \right\| \\ &\leq \int_{q_1}^{q_2} \|h^{(l)}(x_{[t+q]}, s_{[t+q]}, u_{[t+q]}^{(l)})\| dq \\ &\leq \int_{q_1}^{q_2} K^{(l)}(1 + \|x_{[t+q]}\|) dq \\ &\leq C^{(l)}(q_2 - q_1), \end{aligned}$$

where $C^{(l)} := K^{(l)}(1 + r)$ and $r > 0$ is such that, $\sup_{n \geq 0} \|x_n\| \leq r$. Thus $\{\tilde{x}^{(l)}(\cdot; t)\}_{t \geq 0}$ is an equicontinuous family. Now the claim follows from Arzella-Ascoli theorem. ■

From Lemma 6.3 and Lemma 6.4 we conclude the following for almost every ω .

- (1) The family of functions, $\{\bar{x}(\cdot + t)\}_{t \geq 0}$ **is relatively compact** in $\mathcal{C}([0, \infty), \mathbb{R}^d)$, because if not, there exist $t_n \rightarrow \infty$ such that $\{\bar{x}(\cdot + t_n)\}_{n \geq 0}$ does not have a limit point in $\mathcal{C}([0, \infty), \mathbb{R}^d)$. Then by Lemma 6.3, $\{\tilde{x}^{(l)}(\cdot; t_n)\}_{n \geq 0}$ does not have a limit point in $\mathcal{C}([0, \infty), \mathbb{R}^d)$ which contradicts Lemma 6.4.
- (2) Since $\{\bar{x}(\cdot + t)\}_{t \geq 0}$ is relatively compact in $\mathcal{C}([0, \infty), \mathbb{R}^d)$, by Arzella-Ascoli theorem, we have that for every $T > 0$, $\{\bar{x}(\cdot + t)|_{[0, T]}\}_{t \geq 0}$ is equicontinuous. Set $T = 1$ and fix $\epsilon > 0$. Then for any $t_0 \geq \frac{1}{2}$, there exists $t \geq 0$ such that $t + \frac{1}{2} = t_0$. By equicontinuity of $\{\bar{x}(\cdot + t)|_{[0, T]}\}_{t \geq 0}$, we can obtain a $\delta > 0$ (independent of t and hence t_0) such that, for every t' satisfying $|t' - t_0| < \min\{\delta, \frac{1}{2}\}$, we have $\|\bar{x}(t') - \bar{x}(t_0)\| < \epsilon$. Since ϵ was arbitrary and $\bar{x}(\cdot)|_{[0, \frac{1}{2}]}$ is uniformly continuous we obtain that the function $\bar{x}(\cdot)$ **is uniformly continuous** on $[0, \infty)$.

Proposition 6.5 *For a.e. ω , every limit point $x^*(\cdot)$ of $\{\bar{x}(\cdot + t)\}_{t \geq 0}$ satisfies the following.*

- (i) *For every $l \geq 1$, there exists $\tilde{\gamma}^{(l)} \in \mathcal{M}(\mathcal{S} \times U)$ such that, for every $t \geq 0$,*

$$x^*(t) = x^*(0) + \int_0^t \left[\int_{\mathcal{S} \times U} h^{(l)}(x^*(q), s, u) \tilde{\gamma}^{(l)}(q)(ds, du) \right] dq.$$

- (ii) *For every $l \geq 1$, $\tilde{\gamma}^{(l)}$ as in part (i) of this lemma is such that, for almost every $t \geq 0$,*

$$\Lambda(\tilde{\gamma}^{(l)})(t) \in D(x^*(t)).$$

- (iii) *$x^*(\cdot)$ is absolutely continuous and for almost every $t \geq 0$,*

$$\frac{dx^*(t)}{dt} \in \hat{H}(x^*(t)).$$

Proof Fix $\omega \in \Omega_{a,s}$, and let $t_n \rightarrow \infty$, such that $\bar{x}(\cdot + t_n) \rightarrow x^*(\cdot)$ in $\mathcal{C}([0, \infty), \mathbb{R}^d)$.

- (i) Fix $l \geq 1$. Consider the $\mathcal{M}(\mathcal{S} \times U)$ valued sequence $\{\gamma^{(l)}(\cdot + t_n)\}_{n \geq 1}$. Since $\mathcal{M}(\mathcal{S} \times U)$ is a compact metric space, there exists a subsequence of the above that converges. Set $\tilde{\gamma}^{(l)}(\cdot)$ to be some limit point and w.l.o.g. assume $\{\gamma^{(l)}(\cdot + t_n)\}_{n \geq 1}$ converges to $\tilde{\gamma}^{(l)}(\cdot)$. Since the sequence $\{\bar{x}(\cdot + t_n)\}_{n \geq 1}$ converges to $x^*(\cdot)$, by Lemma 6.3, we have that $\{\tilde{x}^{(l)}(\cdot; t_n)\}_{n \geq 1}$ also converges to $x^*(\cdot)$ in $\mathcal{C}([0, \infty), \mathbb{R}^d)$. For every $n \geq 1$, by definition of $\tilde{x}^{(l)}(\cdot; t_n)$ we have that for every $t \geq 0$,

$$\begin{aligned} \tilde{x}^{(l)}(t; t_n) &= \bar{x}(t_n) + \int_0^t y^{(l)}(q) dq \\ &= \bar{x}(t_n) + \int_0^t h^{(l)}(x_{[t_n+q]}, s_{[t_n+q]}, u_{[t_n+q]}^{(l)}) dq. \end{aligned}$$

By definition of $\gamma^{(l)}(\cdot)$ (see (10) and recall that $\gamma^{(l)}(\cdot) = \Gamma^{(l)}(\omega, \cdot)$) for every $n \geq 1$ and for every $t \geq 0$ we can write the above as,

$$\tilde{x}^{(l)}(t; t_n) = \bar{x}(t_n) + \int_0^t \left[\int_{\mathcal{S} \times U} h^{(l)}(x_{[t_n+q]}, s, u) \gamma^{(l)}(q + t_n)(ds, du) \right] dq.$$

Therefore, for every $t \geq 0$,

$$\begin{aligned} \lim_{n \rightarrow \infty} [\tilde{x}^{(l)}(t; t_n) - \bar{x}(t_n)] &= \lim_{n \rightarrow \infty} \int_0^t \left[\int_{\mathcal{S} \times U} h^{(l)}(x_{[t_n+q]}, s, u) \gamma^{(l)}(q + t_n)(ds, du) \right] dq. \\ x^*(t) - x^*(0) &= \lim_{n \rightarrow \infty} \int_0^t \left[\int_{\mathcal{S} \times U} h^{(l)}(x_{[t_n+q]}, s, u) \gamma^{(l)}(q + t_n)(ds, du) \right] dq. \end{aligned} \quad (12)$$

Since $\gamma^{(l)}(\cdot + t_n) \rightarrow \tilde{\gamma}^{(l)}(\cdot)$ and by our choice of the topology for $\mathcal{M}(\mathcal{S} \times U)$, we have,

$$\int_0^t \left[\int_{\mathcal{S} \times U} \tilde{f}(q, s, u) \gamma^{(l)}(q + t_n)(ds, du) \right] dq - \int_0^t \left[\int_{\mathcal{S} \times U} \tilde{f}(q, s, u) \tilde{\gamma}^{(l)}(q)(ds, du) \right] dq \rightarrow 0,$$

for all bounded continuous $\tilde{f} : [0, t] \times \mathcal{S} \times U \rightarrow \mathbb{R}$ of the form,

$$\tilde{f}(q, s, u) = \sum_{m=1}^N a_m g_m(q) f_m(s, u),$$

for some $N \geq 1$, scalars a_m and bounded continuous functions g_m, f_m on $[0, t], \mathcal{S} \times U$ respectively, for $1 \leq m \leq N$. By the Stone-Weierstrass theorem, such functions can uniformly approximate any function in $\mathcal{C}([0, t] \times \mathcal{S} \times U, \mathbb{R})$. Thus the above convergence holds true for all real valued continuous functions on $[0, t] \times \mathcal{S} \times U$, implying that $t^{-1} \gamma^{(l)}(q + t_n)(ds, du) dq \rightarrow t^{-1} \tilde{\gamma}^{(l)}(q)(ds, du) dq$ in $\mathcal{P}([0, t] \times \mathcal{S} \times U)$. Thus,

$$\left\| \int_0^t \left[\int_{\mathcal{S} \times U} h^{(l)}(x^*(q), s, u) \gamma^{(l)}(q + t_n)(ds, du) \right] dq - \int_0^t \left[\int_{\mathcal{S} \times U} h^{(l)}(x^*(q), s, u) \tilde{\gamma}^{(l)}(q)(ds, du) \right] dq \right\| \rightarrow 0 \quad (13)$$

as $n \rightarrow \infty$. Since $\{\bar{x}(\cdot + t_n)|_{[0, t]}\}_{n \geq 1}$ converges uniformly to $x^*(\cdot)|_{[0, t]}$ we have that, the function $q \rightarrow x_{[t_n+q]}$ converges uniformly to $x^*(\cdot)|_{[0, t]}$ on $[0, t]$. Using the above and by Lemma 6.2, we have that for every $\epsilon > 0$, there exists N (depending on ϵ) such that, for every $n \geq N$, for every $q \in [0, t]$, we have,

$$\left\| \int_{\mathcal{S} \times U} h^{(l)}(x_{[t_n+q]}, s, u) \gamma^{(l)}(q + t_n)(ds, du) - \int_{\mathcal{S} \times U} h^{(l)}(x^*(q), s, u) \gamma^{(l)}(q + t_n)(ds, du) \right\| < \epsilon. \quad (14)$$

Now,

$$\begin{aligned} & \left\| \int_0^t \left[\int_{\mathcal{S} \times U} h^{(l)}(x_{[t_n+q]}, s, u) \gamma^{(l)}(q + t_n)(ds, du) \right] dq - \int_0^t \left[\int_{\mathcal{S} \times U} h^{(l)}(x^*(q), s, u) \tilde{\gamma}^{(l)}(q)(ds, du) \right] dq \right\| \\ & \leq \\ & \left\| \int_0^t \int_{\mathcal{S} \times U} h^{(l)}(x_{[t_n+q]}, s, u) \gamma^{(l)}(q + t_n)(ds, du) dq - \int_0^t \int_{\mathcal{S} \times U} h^{(l)}(x^*(q), s, u) \gamma^{(l)}(q + t_n)(ds, du) dq \right\| \\ & + \\ & \left\| \int_0^t \left[\int_{\mathcal{S} \times U} h^{(l)}(x^*(q), s, u) \gamma^{(l)}(q + t_n)(ds, du) \right] dq - \int_0^t \left[\int_{\mathcal{S} \times U} h^{(l)}(x^*(q), s, u) \tilde{\gamma}^{(l)}(q)(ds, du) \right] dq \right\|. \end{aligned}$$

Taking limit on both sides as $n \rightarrow \infty$ in the above equation and using equations (13) and (14) we obtain,

$$\lim_{n \rightarrow \infty} \left\| \int_0^t \left[\int_{\mathcal{S} \times U} h^{(l)}(x_{[t_n+q]}, s, u) \gamma^{(l)}(q + t_n)(ds, du) \right] dq - \int_0^t \left[\int_{\mathcal{S} \times U} h^{(l)}(x^*(q), s, u) \tilde{\gamma}^{(l)}(q)(ds, du) \right] dq \right\| \leq \epsilon t,$$

for every $\epsilon > 0$. Therefore, for every $t \geq 0$,

$$\lim_{n \rightarrow \infty} \int_0^t \left[\int_{\mathcal{S} \times U} h^{(l)}(x_{[t_n+q]}, s, u) \gamma^{(l)}(q + t_n)(ds, du) \right] dq = \int_0^t \left[\int_{\mathcal{S} \times U} h^{(l)}(x^*(q), s, u) \tilde{\gamma}^{(l)}(q)(ds, du) \right] dq.$$

Substituting the above limit in equation (12), we get, for every $t \geq 0$,

$$x^*(t) - x^*(0) = \int_0^t \left[\int_{\mathcal{S} \times U} h^{(l)}(x^*(q), s, u) \tilde{\gamma}^{(l)}(q)(ds, du) \right] dq.$$

- (ii) The proof of this part is similar to the proof of Lemma 6, chapter 6.3 of [10]. We shall present a proof here for the sake of completeness. Let $\{f_i\}$ be a countable set of real valued

continuous functions on \mathcal{S} that is a convergence determining class for $\mathcal{P}(\mathcal{S})$. By replacing each f_i by $a_i f_i + b_i$ for suitable scalars $a_i, b_i > 0$, we may suppose that $0 \leq f_i(\cdot) \leq 1$ for all i . For each i ,

$$\zeta_n^i := \sum_{k=0}^{n-1} a(k)(f_i(S_{k+1}) - \int_{\mathcal{S}} f_i(s') \Pi(X_k, S_k)(ds')),$$

is a square integrable zero mean martingale w.r.t. the filtration $\{\mathcal{F}_n := \sigma(S_k, X_k : 0 \leq k \leq n-1)\}_{n \geq 1}$ and for almost every ω , $\sum_{n=1}^{\infty} \mathbb{E}[(\zeta_{n+1}^i - \zeta_n^i)^2 | \mathcal{F}_n] \leq 2 \sum_{n=0}^{\infty} a(n)^2 < \infty$. Hence by martingale convergence theorem (see Appendix C, Theorem 11 in [10]), for almost every ω , $\{\zeta_n^i\}_{n \geq 1}$ converges. Let $\Omega_m := \left\{ \omega \in \Omega : \forall i, \{\zeta_n^i\}_{n \geq 1} \text{ converges} \right\}$. Then $\mathbb{P}(\Omega_m) = 1$. Define,

$$\Omega^* := \Omega_m \cap \Omega_{a,s} \quad (15)$$

and clearly $\mathbb{P}(\Omega^*) = 1$. Recall, that for every $T \geq 0$ and $n \geq 0$, $\tau(n, T) := \min \{k \geq n : t(k) \geq t(n) + T\}$. Then for every $\omega \in \Omega^*$, for every i , for every $T > 0$, as $n \rightarrow \infty$,

$$\sum_{k=n}^{\tau(n,T)} a(k)(f_i(s_{k+1}) - \int_{\mathcal{S}} f_i(s') \Pi(x_k, s_k)(ds')) \rightarrow 0.$$

By the choice of $\{f_i\}_{i \geq 1}$ and the fact that $\{a(n)\}_{n \geq 0}$ are non-increasing (see assumption (A3)(i)) we get, that for every $\omega \in \Omega^*$, for every i , for every $T > 0$, as $n \rightarrow \infty$,

$$\left| \sum_{k=n}^{\tau(n,T)} (a(k) - a(k+1)) f_i(s_{k+1}) \right| \leq a(n) - a(\tau(n, T) + 1) \rightarrow 0.$$

Thus, for every $\omega \in \Omega^*$, for every i , for every $T > 0$, as $n \rightarrow \infty$,

$$\sum_{k=n}^{\tau(n,T)} a(k)(f_i(s_k) - \int_{\mathcal{S}} f_i(s') \Pi(x_k, s_k)(ds')) \rightarrow 0. \quad (16)$$

Fix $l \geq 1$. Define for every ω , $\mu^{(l)}(\cdot) := \Lambda(\gamma^{(l)}) \in \mathcal{M}(\mathcal{S})$. Then by definition of $\gamma^{(l)}(\cdot)$ (see equation (10) and recall $\gamma^{(l)}(\cdot) := \Gamma^{(l)}(\omega, \cdot)$) and $\Lambda(\cdot)$ (see Lemma 5.2) we have that for every ω , for every $t > 0$, $\mu^{(l)}(t) = \delta_{s_n}$ where n is such that, $t \in [t(n), t(n+1))$. Using the definition of $\mu^{(l)}(\cdot)$ in equation (16), we get that for every $\omega \in \Omega^*$, for every i , for every $T > 0$, as $n \rightarrow \infty$,

$$\int_0^{t(\tau(n,T)) - t(n)} \int_{\mathcal{S}} \left[f_i(s) - \int_{\mathcal{S}} f_i(s') \Pi(x_{[t(n)+q]}, s)(ds') \right] \mu^{(l)}(q + t(n))(ds) dq \rightarrow 0,$$

where, for every $t \geq 0$, $[t] := \max \{n \geq 0 : t(n) \leq t\}$. From the above it can easily be shown that for every $\omega \in \Omega^*$, for every i , for every $T > 0$,

$$\lim_{t \rightarrow \infty} \int_0^T \int_{\mathcal{S}} \left[f_i(s) - \int_{\mathcal{S}} f_i(s') \Pi(x_{[t+q]}, s)(ds') \right] \mu^{(l)}(q + t)(ds) dq = 0.$$

By assumption (A2), $\Pi(\cdot)|_{2rU \times \mathcal{S}}$ is uniformly continuous, where $r = \sup_{n \geq 0} \|x_n\|$. Thus the function $(x, s) \rightarrow f_i(s) - \int_{\mathcal{S}} f_i(s') \Pi(x, s)(ds')$ is uniformly continuous on $2rU \times \mathcal{S}$ for every i . Using the above, the fact that $\lim_{n \rightarrow \infty} (t(n+1) - t(n)) = 0$ and uniform continuity of $\bar{x}(\cdot)$ we get that, for every $\omega \in \Omega^*$, for every i , for every $T > 0$,

$$\lim_{t \rightarrow \infty} \int_0^T \int_{\mathcal{S}} \left[f_i(s) - \int_{\mathcal{S}} f_i(s') \Pi(\bar{x}(t+q), s)(ds') \right] \mu^{(l)}(q + t)(ds) dq = 0.$$

Fix $\omega \in \Omega^*$. From part (i) of this lemma we have a sequence $t_n \rightarrow \infty$ such that $\{\bar{x}(\cdot + t_n)\}_{n \geq 1}$ converges to $x^*(\cdot)$ in $\mathcal{C}([0, \infty), \mathbb{R}^d)$ and $\{\gamma^{(l)}(\cdot + t_n)\}_{n \geq 1}$ converges to $\tilde{\gamma}^{(l)}(\cdot)$ in $\mathcal{M}(\mathcal{S} \times U)$. By continuity of the map $\Lambda(\cdot)$ (see Lemma 5.2), we have that $\mu^{(l)}(\cdot + t_n) \rightarrow \tilde{\mu}^{(l)}(\cdot) = \Lambda(\tilde{\gamma}^{(l)})$ in $\mathcal{M}(\mathcal{S})$. Using the convergence above and the fact that the family of functions, $\{x \in 2rU \rightarrow \int_{\mathcal{S}} [f_i(s) - \int_{\mathcal{S}} f_i(s') \Pi(x, s)(ds')] \nu(ds) : \nu \in \mathcal{P}(\mathcal{S})\}$ is equicontinuous (which can be shown by arguments similar to Lemma 6.2) we get that for every i , for every $T > 0$,

$$\int_0^T \int_{\mathcal{S}} \left[f_i(s) - \int_{\mathcal{S}} f_i(s') \Pi(x^*(q), s)(ds') \right] \tilde{\mu}^{(l)}(q)(ds) dq = 0.$$

An application of Lebesgue's theorem (see chapter 11.1.3 in [10]), we get that for almost every $t \geq 0$, for every i ,

$$\int_{\mathcal{S}} \left[f_i(s) - \int_{\mathcal{S}} f_i(s') \Pi(x^*(t), s)(ds') \right] \tilde{\mu}^{(l)}(t)(ds) = 0.$$

By our choice of $\{f_i\}_{i \geq 1}$, we get that for almost every $t \geq 0$,

$$\tilde{\mu}^{(l)}(t)(ds) = \int_{\mathcal{S}} \Pi(x^*(t), s')(ds) \tilde{\mu}^{(l)}(t)(ds').$$

Therefore for almost every $t \geq 0$, $\tilde{\mu}^{(l)}(t) = \Lambda(\tilde{\gamma}^{(l)})(t) \in D(x^*(t))$.

(iii) Fix $l \geq 1$. From part (i) of this lemma we have that for every $t \geq 0$,

$$x^*(t) = x^*(0) + \int_0^t \left[\int_{\mathcal{S} \times U} h^{(l)}(x^*(q), s, u) \tilde{\gamma}^{(l)}(q)(ds, du) \right] dq.$$

Clearly $x^*(\cdot)$ is absolutely continuous and for almost every $t \geq 0$,

$$\frac{dx^*(t)}{dt} = \int_{\mathcal{S} \times U} h^{(l)}(x^*(t), s, u) \tilde{\gamma}^{(l)}(t)(ds, du). \quad (17)$$

By part (ii) of this lemma we have that for almost every $t \geq 0$, $\Lambda(\tilde{\gamma}^{(l)})(t) \in D(x^*(t))$. By definition of the map $\Lambda(\cdot)$ (see Lemma 5.2), we have that for almost every $t \geq 0$, $\tilde{\gamma}_S^{(l)}(t) \in D(x^*(t))$. By Lemma 4.2 and by definition of $\hat{H}^{(l)}$ (see equation (6)) we have that for almost every $t \geq 0$,

$$\int_{\mathcal{S} \times U} h^{(l)}(x^*(t), s, u) \tilde{\gamma}^{(l)}(t)(ds, du) \in \cup_{\mu \in D(x^*(t))} \int_{\mathcal{S}} H_{x^*(t)}^{(l)}(s) \mu(ds) = \hat{H}^{(l)}(x^*(t)).$$

Using the above in equation (17) we obtain that for almost every $t \geq 0$,

$$\frac{dx^*(t)}{dt} \in \hat{H}^{(l)}(x^*(t)).$$

Since $l \geq 1$ that was fixed was arbitrary, the above holds for every $l \geq 1$. Therefore, for almost every $t \geq 0$,

$$\frac{dx^*(t)}{dt} \in \cap_{l \geq 1} \hat{H}^{(l)}(x^*(t)) = \hat{H}(x^*(t)),$$

where the equality follows from Lemma 4.4(iii). \blacksquare

Before we proceed further we shall briefly recall the definition of asymptotic pseudotrajectories (APT) for set-valued dynamics introduced in [1]. The translation flow $\Theta : \mathcal{C}(\mathbb{R}, \mathbb{R}^d) \times \mathbb{R} \rightarrow \mathcal{C}(\mathbb{R}, \mathbb{R}^d)$ is the flow defined by,

$$\Theta^t(\mathbf{x})(q) := \mathbf{x}(q + t).$$

For every $\omega \in \Omega$, extend $\bar{x}(\cdot)$ to \mathbb{R} by letting $\bar{x}(t) = \bar{x}(0)$ for $t < 0$. Then $\bar{x}(\cdot)$ is an APT for the flow of DI (8) if,

$$\lim_{t \rightarrow \infty} \mathbf{D}(\Theta^t(\bar{x}), \Sigma) = 0,$$

where, $\Sigma := \cup_{x \in \mathbb{R}^d} \Sigma(x)$ denotes the set of all solutions of DI (8).

In what follows we fix $\omega \in \Omega^*$ and let $\bar{x}(\cdot)$ denote the extension to \mathbb{R} as defined above. By uniform continuity of $\bar{x}(\cdot)$, we have that the family $\{\Theta^t(\bar{x})\}_{t \geq 0}$ is equicontinuous and by assumption (A5) is pointwise bounded. Hence $\{\Theta^t(\bar{x})\}_{t \geq 0}$ is relatively compact in $\mathcal{C}(\mathbb{R}, \mathbb{R}^d)$. Let $x^*(\cdot)$ be a limit point of $\{\Theta^t(\bar{x})\}_{t \geq 0}$. Then by Proposition 6.5(iii), we have that $x^*(\cdot)|_{[0, \infty)}$ is a solution on $[0, \infty)$ of DI (8). Usually the negative time argument is omitted since it follows from the positive time argument as follows:

Fix $T > 0$. Since $x^*(\cdot)$ is a limit point of $\{\Theta^t(\bar{x})\}_{t \geq 0}$, there exists $t_n \rightarrow \infty$ such that, $\{\Theta^{t_n}(\bar{x})\}_{n \geq 1}$ converges to $x^*(\cdot)$ in $\mathcal{C}(\mathbb{R}, \mathbb{R}^d)$. Then $\{\Theta^{t_n - T}(\bar{x})\}_{n \geq 1}$ converges to $x^*(\cdot - T)$. By Proposition 6.5(iii), $x^*(\cdot - T)|_{[0, \infty)}$ is a solution of DI (8). Therefore $x^*(\cdot)|_{[-T, 0]}$ is absolutely continuous and for almost every $t \in [-T, 0]$,

$$\frac{dx^*(t)}{dt} \in \hat{H}(x^*(t)).$$

Since $T > 0$, is arbitrary, we have that $x^*(\cdot)|_{(-\infty, 0]}$ is a solution on $(-\infty, 0]$ of DI (8).

Therefore every limit point of $\{\Theta^t(\bar{x})\}_{t \geq 0}$ is in Σ , the set of solutions of DI (8). Then by Theorem 4.1 in [1] we get the following result.

Theorem 6.6 [APT] *Under assumptions (A1) – (A5), for almost every ω , the linearly interpolated trajectory of recursion (4), $\bar{x}(\cdot)$, is an asymptotic pseudotrajectory of DI (8).*

6.3 Characterization of limit sets

For every ω , the limit set of recursion (4), denoted by $L(\bar{x})$ is defined as,

$$L(\bar{x}) := \cap_{t \geq 0} \overline{\{\bar{x}(q + t) : q \geq 0\}}. \quad (18)$$

As a consequence of Theorem 6.6 we will be able to characterize the limit set of recursion (4) in terms of the dynamics induced by \hat{H} . The notions of invariance, internal chain transitivity, attracting sets, basin of attraction and attractors are taken from [1]. We shall state here definitions of a few of the notions mentioned above for the sake of completeness.

The **flow** of DI(8) is given by the set valued map $\Phi : \mathbb{R}^d \times \mathbb{R} \rightarrow \{\text{subsets of } \mathbb{R}^d\}$, where for every $(x, t) \in \mathbb{R}^d \times \mathbb{R}$, $\Phi(x, t) := \{\mathbf{x}(t) : \mathbf{x} \in \Sigma(x)\}$. For any $C \subseteq \mathbb{R}^d$, let $\omega_\Phi(C) := \cap_{t \geq 0} \overline{\Phi_{[t, \infty)}(C)}$, where $\Phi_{[t, \infty)}(C) := \cup_{(x, \tau) \in C \times [t, \infty)} \Phi(x, \tau)$. For any $A \subseteq \mathbb{R}^d$, its **basin of attraction** denoted by $B(A)$ is defined as, $B(A) := \{x \in \mathbb{R}^d : \omega_\Phi(\{x\}) \subseteq A\}$.

A set $A \subseteq \mathbb{R}^d$ is said to be **invariant** for DI(8) if for all $x \in A$, there exists $\mathbf{x} \in \Sigma(x)$ such that for every $t \in \mathbb{R}$, $\mathbf{x}(t) \in A$.

A compact set $A \subseteq \mathbb{R}^d$ is an **attracting set** for the flow of DI(8) if there exists a neighborhood O of A , with the property that for every $\epsilon > 0$ there exists $t_\epsilon > 0$ such that for every $t \geq t_\epsilon$, $\Phi_t(O) \subseteq N^\epsilon(A)$, where $N^\epsilon(A)$ stands for the ϵ -neighborhood of A .

A compact set $A \subseteq \mathbb{R}^d$ is an **attractor** for the flow of DI(8) if it is an attracting set and is invariant. Further if the basin of attraction of attractor A is the whole of \mathbb{R}^d , that is $B(A) = \mathbb{R}^d$, then A is a **global attractor**.

Given a set $A \subseteq \mathbb{R}^d$ and $x, y \in A$, for any $\epsilon > 0$ and $T > 0$ there exists an (ϵ, T) chain from x to y for DI(8) if there exists an integer $n \in \mathbb{N}$, solutions $\mathbf{x}_1, \dots, \mathbf{x}_n$ to DI(8) and real numbers t_1, \dots, t_n greater than T such that

- for all $i \in \{1, \dots, n\}$ and for all $q \in [0, t_i]$, $\mathbf{x}_i(q) \in A$,

- for all $i \in \{1, \dots, n\}$, $\|\mathbf{x}_i(t_i) - \mathbf{x}_{i+1}(0)\| \leq \epsilon$,
- $\|\mathbf{x}_1(0) - x\| \leq \epsilon$ and $\|\mathbf{x}_n(t_n) - y\| \leq \epsilon$.

A compact set $A \subseteq \mathbb{R}^d$ is said to be **internally chain transitive** if for every $x, y \in A$, for every $\epsilon > 0$ and for every $T > 0$, there exists (ϵ, T) chain from x to y for the DI(8).

Theorem 6.7 [Limit set] *Under assumptions (A1) – (A5), for almost every ω , the following hold.*

- (i) $L(\bar{x})$ is a non-empty, compact subset of \mathbb{R}^d and is internally chain transitive.
- (ii) If $A \subseteq \mathbb{R}^d$ is an attracting set for the flow of DI (8) with a basin of attraction $B(A)$, such that $L(\bar{x}) \cap B(A) \neq \emptyset$, then $L(\bar{x}) \subseteq A$.
- (iii) If $A \subseteq \mathbb{R}^d$ is a global attractor for the flow of DI (8), then $L(\bar{x}) \subseteq A$.
- (iv) If $A = \{x^*\}$ is a global attractor for the flow of DI (8), then $\{x_n\}_{n \geq 0}$ converges to x^* .

Proof :

- (i) Follows from Theorem 6.6 above and Theorem 4.3 of [1].
- (ii) From part (i) of this theorem we have that $L(\bar{x})$ is internally chain transitive. Now the claim follows from Theorem 3.23 of [1].
- (iii) Follows from part (i) of this theorem and Corollary 3.24 of [1].
- (iv) Follows from part (iii) of this theorem with $A = x^*$. ■

7 Iterate-independent and controlled Markov noise cases

In this section we shall introduce two variants of assumption (A2) under which a similar analysis as presented in the previous section can be carried out (assumptions (A1), (A3) – (A5) are assumed to hold). The first case is of iterate-independent Markov noise i.e., where the transition probabilities do not depend on the iterate x while second case is of controlled Markov noise i.e., the so-called noise process is not Markov by itself, but its lack of Markov property comes through its dependence on some possibly imperfectly known time varying process which is viewed as a control for analysis purposes. Throughout this section we assume that (A1), (A3) – (A5) hold.

7.1 Iterate-independent Markov noise

First present the iterate-independent Markov noise assumption.

(A2)' $\{S_n\}_{n \geq 0}$ is a sequence of \mathcal{S} valued random variables on Ω such that, for every $n \geq 0$, for every $A \in \mathcal{B}(\mathcal{S})$, $\mathbb{P}(S_{n+1} \in A | S_n, X_n, n \leq n) = \mathbb{P}(S_{n+1} \in A | S_n) = \Pi(S_n)(A)$ a.s. where $\Pi : \mathcal{S} \rightarrow \mathcal{P}(\mathcal{S})$ is continuous.

Assumption (A2)' and the fact that the state space of Markov noise is a compact metric space gives us that the Markov chain defined by the transition kernel, $\Pi(\cdot)(\cdot)$, admits a stationary distribution ($\mu \in \mathcal{P}(\mathcal{S})$ is stationary for the Markov chain defined by the transition kernel $\Pi(\cdot)(\cdot)$ if, for every $A \in \mathcal{B}(\mathcal{S})$, $\mu(A) = \int_{\mathcal{S}} \Pi(s)(A) \mu(ds)$). Let $D \subseteq \mathcal{P}(\mathcal{S})$ denote the set of stationary distributions of the Markov chain defined by the transition kernel $\Pi(\cdot)(\cdot)$. It is easy to show that D is a convex and compact subset of $\mathcal{P}(\mathcal{S})$. Define the set valued map, $\hat{H}_i : \mathbb{R}^d \rightarrow \{\text{subsets of } \mathbb{R}^d\}$, such that, for every $x \in \mathbb{R}^d$,

$$\hat{H}_i(x) := \cup_{\mu \in D} \int_{\mathcal{S}} H_x(s) \mu(ds),$$

where for every $x \in \mathbb{R}^d$, the set valued map H_x is as defined in section 4. Similar to Lemma 4.5, it can be shown that the set valued map \hat{H} defined above is a Marchaud map. By the same set of arguments as in section 6, we obtain the following result.

Theorem 7.1 [APT for iterate-independent Markov noise] *Under assumptions (A1), (A2)', (A3)–(A5), for almost every ω , the linearly interpolated trajectory of recursion (4), is an asymptotic pseudotrajectory for the flow of the DI,*

$$\frac{dx}{dt} \in \hat{H}_i(x). \quad (19)$$

As a consequence of the above theorem, a limit set theorem similar to Theorem 6.7 can be established which under assumptions (A1), (A2)', (A3) – (A5) characterizes the limit set of recursion (4) in terms of the dynamics of DI(19).

7.2 Controlled Markov noise

Let $\{Z_n\}_{n \geq 0}$ be a sequence of random variables (control sequence) on Ω , taking values in a compact metric space, \mathcal{Z} . The controlled Markov noise assumption is as follows.

(A2)'' $\{S_n\}_{n \geq 0}$ is a sequence of \mathcal{S} valued random variables on Ω such that, for every $n \geq 0$, for every $A \in \mathcal{B}(\mathcal{S})$, $\mathbb{P}(S_{n+1} \in A | S_n, Z_n, X_n) = \mathbb{P}(S_{n+1} \in A | S_n, Z_n, X_n) = \Pi(X_n, S_n, Z_n)(A)$ a.s. with $\Pi : \mathcal{S} \rightarrow \mathcal{P}(\mathcal{S})$, satisfying,

- (i) Π is continuous.
- (ii) $\{Z_n\}_{n \geq 0}$ is a stationary randomized control, i.e., for every $n \geq 0$, for every $A \subseteq \mathcal{B}(\mathcal{Z})$, $\mathbb{P}(Z_n \in A | X_n, S_n, Z_{n-1}, m \leq n) = \mathbb{P}(Z_n \in A | S_n) = \phi(S_n)(A)$ a.s. with $\phi : \mathcal{S} \rightarrow \mathcal{P}(\mathcal{Z})$, measurable.

Let $\Pi_\phi : \mathbb{R}^d \times \mathcal{S} \rightarrow \mathcal{P}(\mathcal{S})$ be such that for every $x \in \mathbb{R}^d$, for every $s \in \mathcal{S}$, $\Pi_\phi(x, s)(ds') = \int_{\mathcal{Z}} \Pi(x, z, s)(ds') \phi(s)(dz)$.

- (iii) For every $x \in \mathbb{R}^d$ the Markov chain with state space \mathcal{S} defined by the transition kernel, $\Pi_\phi(x, \cdot)(\cdot)$ admits a stationary distribution ($\mu \in \mathcal{P}(\mathcal{S})$ is stationary for the Markov chain defined by the transition kernel $\Pi_\phi(x, \cdot)(\cdot)$ if, for every $A \in \mathcal{B}(\mathcal{S})$, $\mu(A) = \int_{\mathcal{S}} \Pi_\phi(x, s)(A) \mu(ds)$).

For every $x \in \mathbb{R}^d$, let $D_\phi(x) \subseteq \mathcal{P}(\mathcal{S})$ denote the set of stationary distributions of the Markov chain defined by the transition kernel, $\Pi_\phi(x, \cdot)(\cdot)$. The convexity, compactness of $D_\phi(x)$ and the closed graph property of the set valued map $x \rightarrow D_\phi(x)$ follow from arguments similar to those in page 69 of [10]. Define the set valued map, $\hat{H}_\phi : \mathbb{R}^d \rightarrow \{\text{subsets of } \mathbb{R}^d\}$ such that for every $x \in \mathbb{R}^d$,

$$\hat{H}_\phi(x) := \bigcup_{\mu \in D_\phi(x)} \int_{\mathcal{S}} H_x(s) \mu(ds),$$

where for every $x \in \mathbb{R}^d$, the set valued map H_x is as defined in section 4. By a proof similar to Lemma 4.5, one can show that the map \hat{H}_ϕ is a Marchaud map. Then by a similar set of arguments as in section 6, one obtains the following result.

Theorem 7.2 [APT for controlled Markov noise] *Under assumptions (A1), (A2)'', (A3) – (A5), for almost every ω , the linearly interpolated trajectory of recursion (4) is an asymptotic pseudotrajectory for the flow of the DI,*

$$\frac{dx}{dt} \in \hat{H}_\phi(x). \quad (20)$$

As a consequence of the above theorem, a limit set theorem similar to Theorem 6.7 can be established which under assumptions (A1), (A2)'', (A3) – (A5) characterizes the limit set of recursion (4) in terms of the dynamics of DI(20).

8 Applications

In this section we consider four applications of the foregoing. These applications presented are natural extensions of the ones presented in chapter 5.3 of [10] to the case with Markov noise. All the applications presented in this section are stated with Markov noise terms $\{S_n\}_{n \geq 0}$ satisfying assumption (A2). One may also consider the applications presented here with the Markov noise terms satisfying assumptions (A2)' or (A2)'' as well.

8.1 Controlled stochastic approximation

Consider the iteration

$$X_{n+1} - X_n - a(n)M_{n+1} = a(n)h(X_n, Z_n, S_n), \quad (21)$$

where, $\{Z_n\}_{n \geq 0}$ is a random sequence taking values in a compact metric space \mathcal{Z} and $h : \mathbb{R}^d \times \mathcal{Z} \times \mathcal{S} \rightarrow \mathbb{R}^d$ is continuous and Lipschitz in the first argument uniformly w.r.t. the second and third, i.e., there exists $L > 0$ such that, for every $(z, s) \in \mathcal{Z} \times \mathcal{S}$ and for every $x_1, x_2 \in \mathbb{R}^d$,

$$\| h(x_1, z, s) - h(x_2, z, s) \| \leq L \| x_1 - x_2 \|.$$

$\{Z_n\}_{n \geq 0}$ is viewed here as a control sequence, that is, Z_n is chosen by the agent running the algorithm at time $n \geq 0$ based on the observed history and possibly extraneous independent randomization as is usual in stochastic control problems. It could also be an unknown random process that affects the measurements in addition to the additive noise $\{M_n\}_{n \geq 0}$. We shall assume that assumptions (A2) – (A5) hold.

Define the set valued map $H : \mathbb{R}^d \times \mathcal{S} \rightarrow \{\text{subsets of } \mathbb{R}^d\}$ such that, for every $(x, s) \in \mathbb{R}^d \times \mathcal{S}$,

$$H(x, s) := \bar{co}(\{h(x, z, s) : z \in \mathcal{Z}\}).$$

Then it can be shown that the set valued map H satisfies assumption (A1) (see Lemma 5 in chapter 5.3 of [10]). Then recursion (21) can be written as,

$$X_{n+1} - X_n - a(n)M_{n+1} \in a(n)H(X_n, S_n)$$

and the asymptotic behavior of the same can be analyzed as in section 6 of this paper.

8.2 Stochastic subgradient descent of a parametrized convex function

Consider the function, $J : \mathbb{R}^d \times \mathcal{S} \rightarrow \mathbb{R}$, continuous and for every $s \in \mathcal{S}$, $J(\cdot, s)$ is a convex function. Suppose $J(\cdot, s)$ is not continuously differentiable everywhere, so that $\nabla J(x, s)$ is not defined at all $x \in \mathbb{R}^d$, where ∇ denotes gradient w.r.t. x . A natural generalization of gradient to this non-smooth case is the notion of subdifferential w.r.t. x denoted by $\partial J(x, s)$, defined as the set of all $y \in \mathbb{R}^d$ such that for every $z \in \mathbb{R}^d$,

$$J(z, s) \geq J(x, s) + \langle y, z - x \rangle.$$

Then for every $(x, s) \in \mathbb{R}^d \times \mathcal{S}$, $\partial J(x, s)$ is nonempty, convex and compact (see Proposition 5.4.1 in [18]). From the definition of subdifferential and the continuity of function $J(\cdot)$, it can be easily shown that the set valued map $(x, s) \rightarrow \partial J(x, s)$ has a closed graph.

Define the set valued map $H : \mathbb{R}^d \times \mathcal{S} \rightarrow \{\text{subsets of } \mathbb{R}^d\}$, such that for every $(x, s) \in \mathbb{R}^d \times \mathcal{S}$,

$$H(x, s) = -\partial J(x, s). \quad (22)$$

We further assume that the set valued map H satisfies the linear growth property (that is assumption (A1)(ii)). Then by arguments in the preceding paragraph the set valued map H satisfies assumption (A1).

Consider the recursion (4) with the set valued map defined above. The recursion,

$$X_{n+1} - X_n - a(n)M_{n+1} \in a(n)H(X_n, S_n),$$

where $\{M_n\}_{n \geq 1}$ can be interpreted as subgradient estimation error. One case of interest where assumption (A4) associated with the additive noise terms is satisfied is when M_n are i.i.d. zero mean with finite variance (more generally martingale difference terms satisfying assumption (A3) in chapter 2.1 of [10]).

Suppose the assumptions (A2) – (A5) are satisfied then the asymptotic behavior of the above recursion can be analyzed as in section 6 of this paper. If in addition to assumption (A2), we know that for every $x \in \mathbb{R}^d$, $|D(x)| = 1$ (that is the stationary distribution is unique for every x), then the set valued map associated with the limiting DI, \hat{H} as defined in (5) is such that for every $x \in \mathbb{R}^d$,

$$\hat{H}(x) \subseteq -\partial \mathbb{E}_s[J(\cdot); \mu_x](x),$$

where $\{\mu_x\} = D(x)$, $\mathbb{E}_s[J(\cdot); \mu_x]$ denotes a real valued map on \mathbb{R}^d such that, for every $y \in \mathbb{R}^d$, $\mathbb{E}_s[J(\cdot); \mu_x](y) = \int_{\mathcal{S}} J(y, s) \mu_x(ds)$ (which is clearly a convex function) and $\partial \mathbb{E}_s[J(\cdot); \mu_x](x)$ denotes the subdifferential of the above defined map at x . If in addition we have that the state space of the Markov noise is finite (that is $|\mathcal{S}| < \infty$) then by Proposition 5.4.6 in [18] we have that for every $x \in \mathbb{R}^d$,

$$\hat{H}(x) = -\partial \mathbb{E}_s[J(\cdot); \mu_x](x).$$

As a consequence of the above we have that the global attractor of the DI,

$$\frac{dx}{dt} \in -\partial \mathbb{E}_s[J(\cdot); \mu_x](x),$$

is also a global attractor of the limiting DI (that is DI (8)) and hence by the limit set theorem (that is Theorem 6.7), we can conclude that the iterates $\{X_n\}_{n \geq 0}$ will almost surely converge to such an attractor.

Another important application is the minimization of the convex function obtained by averaging the parameter w.r.t. a particular probability distribution say, $\mu \in \mathcal{P}(\mathcal{S})$. Formally the minimization of function $J_\mu : \mathbb{R}^d \rightarrow \mathbb{R}$, where for every $x \in \mathbb{R}^d$, $J_\mu(x) := \int_{\mathcal{S}} J(x, s) \mu(ds)$. Consider recursion (4), where the set-valued map H is as in equation (22) and Markov noise terms (that is $\{S_n\}_{n \geq 0}$) satisfy assumption (A2) with the additional condition that for every $x \in \mathbb{R}^d$, $D(x) = \{\mu\}/D_\phi(x) = \{\mu\}$. Then under additional assumptions of (A1), (A3) – (A5) we have that the linearly interpolated trajectory of recursion (4) is an APT for the flow of DI (8). The set valued map \hat{H} associated with DI (8), can be shown to satisfy,

$$\hat{H}(x) \subseteq -\partial J_\mu(x),$$

for every $x \in \mathbb{R}^d$. Further if the state space of the Markov noise is finite (that is $|\mathcal{S}| < \infty$) then by Proposition 5.4.6 in [18] we have that for every $x \in \mathbb{R}^d$,

$$\hat{H}(x) = -\partial J_\mu(x).$$

8.3 Approximate drift problem

Consider the continuous function $h : \mathbb{R}^d \times \mathcal{S} \rightarrow \mathbb{R}^d$, satisfying the linear growth property, that is, there exists $K > 0$, such that for every $(x, s) \in \mathbb{R}^d \times \mathcal{S}$, $\|h(x, s)\| \leq K(1 + \|x\|)$. In many applications one wishes to implement the recursion,

$$X_{n+1} - X_n - a(n)M_{n+1} = a(n)h(X_n, S_n),$$

where assumptions (A2)/(A2)'/(A2)'', (A3) – (A5) are assumed to hold. But in practice one has access only to approximate value of the function $h(\cdot)$. Further the magnitude of the error in

approximation of $h(\cdot)$ is only known to be bounded by a constant, say $\epsilon > 0$. Then the recursion implemented becomes,

$$X_{n+1} - X_n - a(n)M_{n+1} = a(n)(h(X_n, S_n) + \eta_n), \quad (23)$$

where η_n represents the error in estimation of $h(\cdot)$ and we have that for every $n \geq 0$, $\|\eta_n\| \leq \epsilon$.

Define the set valued map $H : \mathbb{R}^d \times \mathcal{S} \rightarrow \{\text{subsets of } \mathbb{R}^d\}$ such that for every $(x, s) \in \mathbb{R}^d \times \mathcal{S}$,

$$H(x, s) := h(x, s) + \epsilon U, \quad (24)$$

where U denotes the closed unit ball in \mathbb{R}^d . Then recursion (23) can be written as,

$$X_{n+1} - X_n - a(n)M_{n+1} \in a(n)H(X_n, S_n),$$

where the set valued map H is as defined in equation (24). It can be easily shown that the set valued map H satisfies assumption (A1). Under assumptions (A2) – (A5) one may analyze the asymptotic behavior of recursion (23) as in section 6 of this paper.

8.4 Discontinuous dynamics

Consider the recursion,

$$X_{n+1} - X_n - a(n)M_{n+1} = a(n)h(X_n, S_n), \quad (25)$$

where $h(\cdot)$ is merely measurable, satisfying the linear growth property that is there exists $K > 0$, such that for every $(x, s) \in \mathbb{R}^d \times \mathcal{S}$, $\|h(x, s)\| \leq K(1 + \|x\|)$.

Define the set valued map $H : \mathbb{R}^d \times \mathcal{S} \rightarrow \{\text{subsets of } \mathbb{R}^d\}$ such that for every $(x, s) \in \mathbb{R}^d \times \mathcal{S}$,

$$H(x, s) := \cap_{\epsilon > 0} \bar{co} \left(\{h(y, s') : \|x - y\| < \epsilon, d_{\mathcal{S}}(s, s') < \epsilon\} \right). \quad (26)$$

Then the set valued map defined above can be shown to satisfy assumption (A1). Now the recursion (25) can be written in the form of recursion (4) with the set valued map H as defined in equation (26). Under assumptions (A2) – (A5), one may analyze the asymptotic behavior of recursion (25) as in section 6 of this paper.

9 Conclusions and directions for future work

We have shown that almost surely the linearly interpolated trajectory of recursion (4) is an asymptotic pseudotrajectory for the flow of DI (8). The asymptotic pseudotrajectory result proved in this paper enables one to characterize limit sets of recursion (4) as internally chain transitive sets of the flow of DI (8) which in-turn enables us to guarantee convergence of the iterates to attractors of DI (8) as stated in Theorem 6.7. We have also stated two variants of the Markov noise assumption where the analysis is very similar to the one presented. Finally applications are presented where the recursion studied in this paper naturally appears and in the case of subgradient descent we have also been able to interpret the set-valued map associated with DI (8) as subdifferential of a certain averaged function.

Certain extensions and applications which we wish to consider in future are listed below.

- (1) Stochastic approximation schemes on multiple time scales are extensively studied in literature and find use in several reinforcement learning and optimization applications. For the case without Markov noise, the two time scale stochastic approximation with single valued maps and the case of set-valued maps have already been studied (see chapter 6 in [10] for the single valued case and [19] for the set-valued case). One can extend the analysis presented in this paper to obtain a similar result for two time scale stochastic recursive inclusions with Markov noise.

- (2) Throughout this paper we have assumed that the state space of the Markov noise is a compact metric space. In many practical scenarios, the Markov noise terms take values in a general Polish space. For such a case we believe that the analysis presented in the paper can be extended under certain additional assumptions. The idea would be to embed the Polish space as a dense subset of a compact metric space and under some additional assumptions which guarantee an upper semicontinuous extension of set-valued map H and tightness of some probability measures one might be able to provide an APT argument as in this paper (see [7]).
- (3) Assumption (A5), which ensures that the iterates are stable usually is hard to verify. The Borkar-Meyn theorem (see [11]) which provides a sufficient condition for stability of iterates for stochastic approximation schemes with single valued maps has been extended to the set-valued case without Markov noise (see [13]) and we believe that a similar extension is possible to the set-valued case with Markov noise. Stability results similar to the one in [12] also need to be investigated for possible extension to the set-valued case.
- (4) In many applications one encounters the recursion,

$$X_{n+1} \in \mathbf{P} [X_n + a(n)(H(X_n, S_n) + M_{n+1})],$$

where \mathbf{P} denotes the projection of the set $X_n + a(n)(H(X_n, S_n) + M_{n+1})$ onto a compact (possibly convex) subset of \mathbb{R}^d . Such recursions are of interest also because the assumption of stability (that is (A5)) is naturally satisfied. For the single valued case such recursions are studied in [20] and the extension of the same to the recursion above needs to be established.

- (5) Another variant of Markov noise assumptions studied for the single valued case is given in [6]. These set of assumptions enable one to study behavior of stochastic approximation schemes in the single valued case when the state space is a general Euclidean space. Since in this paper we study the set-valued case by converting it to the single valued setting it would be interesting to see if those assumptions can be used for the set-valued case as in this paper.

A Proof of the continuous embedding lemma (Lemma 3.2)

The proof of this lemma is similar to the proof of Theorem 1, chapter 1, section 13 of [14]. We shall provide a brief outline here for the sake of completeness.

For any $\epsilon > 0$, for every $(x_0, s_0) \in \mathbb{R}^d \times \mathcal{S}$, let $B(\epsilon, x_0, s_0) := \{(x, s) : \|x - x_0\| < \epsilon, d_{\mathcal{S}}(s, s_0) < \epsilon\}$. Let $\{\epsilon_l := \frac{1}{3^l}\}_{l \geq 1}$. Then for every $l \geq 1$, $\mathcal{C}_l := \{B(\epsilon_l, x_0, s_0) : (x_0, s_0) \in \mathbb{R}^d \times \mathcal{S}\}$ is an open covering of $\mathbb{R}^d \times \mathcal{S}$. Since $\mathbb{R}^d \times \mathcal{S}$ is a metric space, it is paracompact (see Defn. 4 and Theorem 1 in chapter 0, section 1 of [14]). Therefore for every $l \geq 1$, there exists a locally finite open refinement of the covering \mathcal{C}_l and let it be denoted by $\tilde{\mathcal{C}}_l := \{C_i^l\}_{i \in I^l}$ where I^l is an arbitrary index set. By Theorem 2, chapter 0, section 1 of [14], there exists a continuous partition of unity, $\{\psi_i^l\}_{i \in I^l}$, subordinated to the covering $\tilde{\mathcal{C}}_l$. Therefore, for every $l \geq 1$, for every $i \in I^l$, there exists (x_i^l, s_i^l) , such that $\text{support}(\psi_i^l) \subseteq C_i^l \subseteq B(\epsilon_l, x_i^l, s_i^l)$. For every $l \geq 1$, for every (x, s) , let $I^l(x, s) := \{i \in I^l : \psi_i^l(x, s) > 0\}$ and by definition of ψ_i^l , we have that $|I^l(x, s)| < \infty$ and $\sum_{i \in I^l(x, s)} \psi_i^l(x, s) = 1$.

For every $l \geq 1$, define the set valued map $H^{(l)} : \mathbb{R}^d \times \mathcal{S} \rightarrow \{\text{subsets of } \mathbb{R}^d\}$, such that for every (x, s) , $H^{(l)}(x, s) := \sum_{i \in I^l(x, s)} \psi_i^l(x, s) A_i^l$, where $A_i^l := \bar{co}(H(B(2\epsilon_l, x_i^l, s_i^l)))$.

Fix $l \geq 1$. For every $i \in I^l$, by assumption (A1)(ii), $\sup_{z \in H(B(2\epsilon_l, x_i^l, s_i^l))} \|z\| \leq K(1 + \|x_i^l\| + 2\epsilon_l)$ from which we can deduce that $\sup_{z \in A_i^l} \|z\| \leq K(1 + \|x_i^l\| + 2\epsilon_l)$. Hence A_i^l is convex and compact. Therefore $H^{(l)}(x, s)$ is a convex combination of compact and convex subsets of \mathbb{R}^d and hence is **convex and compact**.

Fix $l \geq 1$ and (x, s) . Then for every $i \in I^l(x, s)$, $(x, s) \in \text{support}(\psi_i^l) \subseteq C_i^l \subseteq B(\epsilon_l, x_i^l, s_i^l) \subseteq B(2\epsilon_l, x_i^l, s_i^l)$. Therefore for every $i \in I^l(x, s)$, $H(x, s) \subseteq H(B(2\epsilon_l, x_i^l, s_i^l)) \subseteq A_i^l$ and since $H(x, s)$ is convex we have that $H(x, s) = \sum_{i \in I^l(x, s)} \psi_i^l(x, s) H(x, s) \subseteq \sum_{i \in I^l(x, s)} \psi_i^l(x, s) A_i^l = H^{(l)}(x, s)$. Thus $\mathbf{H}(\mathbf{x}, \mathbf{s}) \subseteq \mathbf{H}^{(l)}(\mathbf{x}, \mathbf{s})$.

Fix $l \geq 1$ and (x, s) . Then for every $i \in I^{l+1}(x, s)$, $\|x - x_i^{l+1}\| < \epsilon_{l+1}$ and $d_{\mathcal{S}}(s, s_i^{l+1}) < \epsilon_{l+1}$. Similarly for every $j \in I^l(x, s)$, $\|x - x_j^l\| < \epsilon_l$ and $d_{\mathcal{S}}(s, s_j^l) < \epsilon_l$. Therefore for every $i \in I^{l+1}(x, s)$, for every $j \in I^l(x, s)$, $\|x_i^l - x_i^{l+1}\| < \epsilon_l + \epsilon_{l+1} = \epsilon_l + \frac{\epsilon_l}{3} = \frac{4\epsilon_l}{3}$ and $d_{\mathcal{S}}(s_i^l, s_i^{l+1}) < \frac{4\epsilon_l}{3}$. For every $i \in I^{l+1}(x, s)$, for every $j \in I^l(x, s)$, for every $(x', s') \in B(2\epsilon_{l+1}, x_i^{l+1}, s_i^{l+1})$, $\|x' - x_j^l\| \leq \|x' - x_i^{l+1}\| + \|x_i^{l+1} - x_j^l\| < 2\epsilon_{l+1} + \frac{4\epsilon_l}{3} = \frac{2\epsilon_l}{3} + \frac{4\epsilon_l}{3} = 2\epsilon_l$ and $d_{\mathcal{S}}(s', s_j^l) < 2\epsilon_l$. Thus for every $i \in I^{l+1}(x, s)$, for every $j \in I^l(x, s)$, for every $(x', s') \in B(2\epsilon_{l+1}, x_i^{l+1}, s_i^{l+1})$, $H(x', s') \subseteq H(B(2\epsilon_l, x_j^l, s_j^l)) \subseteq A_j^l$. Therefore for every $i \in I^{l+1}(x, s)$, for every $j \in I^l(x, s)$, $A_i^{l+1} \subseteq A_j^l$ and by using convexity of A_i^{l+1} we get that for every $i \in I^{l+1}(x, s)$, $A_i^{l+1} = \sum_{j \in I^l(x, s)} \psi_j^l(x, s) A_i^{l+1} \subseteq \sum_{j \in I^l(x, s)} \psi_j^l(x, s) A_j^l = H^{(l)}(x, s)$. By using convexity of $H^{(l)}(x, s)$ we get that $H^{(l+1)}(x, s) = \sum_{i \in I^{l+1}(x, s)} \psi_i^{l+1}(x, s) A_i^{l+1} \subseteq \sum_{i \in I^{l+1}(x, s)} \psi_i^{l+1}(x, s) H^{(l)}(x, s) = H^{(l)}(x, s)$. Therefore, $\mathbf{H}^{(l+1)}(\mathbf{x}, \mathbf{s}) \subseteq \mathbf{H}^{(l)}(\mathbf{x}, \mathbf{s})$.

Fix (x, s) . Clearly $H(x, s) \subseteq \cap_{l \geq 1} H^{(l)}(x, s)$. By u.s.c. of H , we have that for every $\epsilon > 0$, there exists $\delta(\epsilon, x, s) > 0$ such that for every (x', s') satisfying $\|x' - x\| < \delta(\epsilon, x, s)$ and $d_{\mathcal{S}}(s', s) < \delta(\epsilon, x, s)$ we have $H(x', s') \subseteq H(x, s) + \epsilon U$. Clearly we can find L such that for every $l \geq L$, $3\epsilon_l < \delta(\epsilon, x, s)$. Therefore for every $\epsilon > 0$, for L as above, we have that for every $l \geq L$, for every $i \in I^l(x, s)$, $H(B(2\epsilon_l, x_i^l, s_i^l)) \subseteq H(x, s) + \epsilon U$. Since $H(x, s) + \epsilon U$ is convex we get, for every $\epsilon > 0$, there exists L such that for every $l \geq L$, for every $i \in I^l(x, s)$, $A_i^l \subseteq H(x, s) + \epsilon U$. Again by convexity of $H(x, s) + \epsilon U$, we have that for every $\epsilon > 0$, there exists L such that $l \geq L$, $H^{(l)}(x, s) = \sum_{i \in I^l(x, s)} \psi_i^l(x, s) A_i^l \subseteq H(x, s) + \epsilon U$. Therefore for every $\epsilon > 0$, $\cap_{l \geq 1} H^{(l)}(x, s) \subseteq H(x, s) + \epsilon U$. Now it follows that $\mathbf{H}(\mathbf{x}, \mathbf{s}) = \cap_{l \geq 1} \mathbf{H}^{(l)}(\mathbf{x}, \mathbf{s})$.

Fix $l \geq 1$ and (x, s) . We have already shown that for every $i \in I^l(x, s)$, $\sup_{z \in A_i^l} \|z\| \leq K(1 + \|x_i^l\| + 2\epsilon_l)$. Using the fact that for every $i \in I^l(x, s)$, $\|x_i^l - x\| < \epsilon_l$ we get that for every $i \in I^l(x, s)$, $\sup_{z \in A_i^l} \|z\| \leq K(1 + \|x\| + 3\epsilon_l)$. Since $\{\psi_i^l(x, s)\}_{i \in I^l(x, s)}$ are convex combining coefficients, we get that $\sup_{\mathbf{z} \in \mathbf{H}^{(l)}(\mathbf{x}, \mathbf{s})} \|\mathbf{z}\| \leq \mathbf{K}^{(l)}(\mathbf{1} + \|\mathbf{x}\|)$, where $K^{(l)} := K + 3\epsilon_l$.

Fix $l \geq 1$ and (x, s) . Since \mathcal{C}_l is a locally finite open covering of $\mathbb{R}^d \times \mathcal{S}$, there exists $\delta > 0$ such that $|I^l(\delta, x, s)| < \infty$ where $I^l(\delta, x, s) := \{i \in I^l : C_i^l \cap B(\delta, x, s) \neq \emptyset\}$. Since for every $i \in I^l$, $\text{support}(\psi_i^l) \subseteq C_i^l$, for every $(x', s') \in B(\delta, x, s)$ we have that $H^{(l)}(x', s') = \sum_{i \in I^l(\delta, x, s)} \psi_i^l(x', s') A_i^l$. Let $z \in H^{(l)}(x, s)$ and $\{(x_n, s_n)\}_{n \geq 1}$ be some sequence converging to (x, s) . Then for every $i \in I^l(\delta, x, s)$ there exists $a_i^l \in A_i^l$ such that $z = \sum_{i \in I^l(\delta, x, s)} \psi_i^l(x, s) a_i^l$ and there exists N , such that for every $n \geq N$, $(x_n, s_n) \in B(\delta, x, s)$. For every $n \leq N$, define z_n to be any value in $H^{(l)}(x_n, s_n)$ and for every $n \geq N$, define $z_n = \sum_{i \in I^l(\delta, x, s)} \psi_i^l(x_n, s_n) a_i^l \in H^{(l)}(x_n, s_n)$ where a_i^l are as defined above. Then clearly for every $n \geq 1$, $z_n \in H^{(l)}(x_n, s_n)$ and $\{z_n\}_{n \geq 1}$ converges to z . Thus $H^{(l)}$ is l.s.c. at (x, s) . Upper simecontinuity of the map $H^{(l)}$ at (x, s) follows from Theorem 1, chapter 1, section 13 of [14]. Since (x, s) is arbitrary the map $H^{(l)}$ is both u.s.c. and l.s.c at every (x, s) and hence is **continuous**.

B Measurability of set-valued map G in Lemma 4.2

Recall that $G : \mathcal{S} \rightarrow \{\text{subsets of } U\}$ is such that for every $s \in \mathcal{S}$, $G(s) := \{u \in U : h^{(l)}(x, s, u) = f(s)\}$. Let $C \subseteq U$ be a closed subset. Then $G^{-1}(C) = \{s \in \mathcal{S} : G(s) \cap C \neq \emptyset\}$. Thus $s \in G^{-1}(C)$ if and only if there exists $u \in C$, such that $h^{(l)}(x, s, u) = f(s)$ or equivalently $s \in G^{-1}(C)$ if and only if $f(s) \in h^{(l)}(x, s, C)$, where $h^{(l)}(x, s, C) := \{h^{(l)}(x, s, u) : u \in C\}$. Since $h_x^{(l)} : \mathcal{S} \times C \rightarrow \mathbb{R}^d$ is continuous, by Proposition 1, chapter 1, section 2 of [14], the set valued map $s \rightarrow h_x^{(l)}(s, C)$ is continuous and its measurability now follows from arguments similar to Lemma 4.1(ii). There-

fore, $G^{-1}(C) = \{s \in \mathcal{S} : f(s) \in h^{(l)}(x, s, C)\} = \{s \in \mathcal{S} : d(f(s), h^{(l)}(x, s, C)) = 0\}$. By Lemma 3.7, $\{s \in \mathcal{S} : d(f(s), h^{(l)}(x, s, C)) = 0\} \in \mathcal{B}(\mathcal{S})$ and hence $G^{-1}(C) \in \mathcal{B}(\mathcal{S})$.

C Proof of uniform convergence in Lemma 4.3(iii)

Since $\{x_{n_k}\}_{k \geq 1}$ converges to x , there exists $r > 0$, such that $\sup_{k \geq 1} \|x_{n_k}\| \leq r$. Fix $\epsilon > 0$. We know that $h^{(l)}(\cdot)|_{2rU \times \mathcal{S} \times U}$ is uniformly continuous and hence there exists $\delta > 0$ (depending on ϵ), such that for every $x' \in \mathbb{R}^d$ satisfying $\|x' - x\| < \delta$, $\sup_{(s,u) \in \mathcal{S} \times U} \|h_x^{(l)}(s, u) - h_{x'}^{(l)}(s, u)\| < \epsilon$. Hence there exists K , such that for every $k \geq K$, $\|x_{n_k} - x\| < \delta$ and therefore for every $k \geq K$, $\sup_{(s,u) \in \mathcal{S} \times U} \|h_{x_{n_k}}^{(l)}(s, u) - h_x^{(l)}(s, u)\| < \epsilon$.

References

- [1] M. Benaïm, J. Hofbauer, and S. Sorin, “Stochastic approximations and differential inclusions,” *SIAM Journal on Control and Optimization*, vol. 44, no. 1, pp. 328–348, 2005.
- [2] M. Benaïm, “A dynamical system approach to stochastic approximations,” *SIAM Journal on Control and Optimization*, vol. 34, no. 2, pp. 437–472, 1996.
- [3] M. Benaïm, “Dynamics of stochastic approximation algorithms,” in *Seminaire de probabilités XXXIII*. Springer, 1999, pp. 1–68.
- [4] M. Metivier and P. Priouret, “Applications of a Kushner and Clark lemma to general classes of stochastic algorithms,” *IEEE Transactions on Information Theory*, vol. 30, no. 2, pp. 140–151, 1984.
- [5] V. Tadić, “Stochastic approximation with random truncations, state-dependent noise and discontinuous dynamics,” *Stochastics: An International Journal of Probability and Stochastic Processes*, vol. 64, no. 3-4, pp. 283–326, 1998.
- [6] G. Fort, E. Moulines, A. Schreck, and M. Vihola, “Convergence of Markovian stochastic approximation with discontinuous dynamics,” *SIAM Journal on Control and Optimization*, vol. 54, no. 2, pp. 866–893, 2016.
- [7] V. S. Borkar, “Stochastic approximation with ‘controlled Markov’ noise,” *Systems & control letters*, vol. 55, no. 2, pp. 139–145, 2006.
- [8] —, *Probability theory: an advanced course*. Springer Science & Business Media, 2012.
- [9] S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.
- [10] V. S. Borkar, *Stochastic approximation : a dynamical systems viewpoint*. Cambridge University Press, 2008.
- [11] V. S. Borkar and S. P. Meyn, “The ODE method for convergence of stochastic approximation and reinforcement learning,” *SIAM Journal on Control and Optimization*, vol. 38, no. 2, pp. 447–469, 2000.
- [12] C. Andrieu, M. Vihola *et al.*, “Markovian stochastic approximation with expanding projections,” *Bernoulli*, vol. 20, no. 2, pp. 545–585, 2014.
- [13] A. Ramaswamy and S. Bhatnagar, “A generalization of the Borkar-Meyn theorem for stochastic recursive inclusions,” *arXiv preprint arXiv:1502.01953*, 2015.

- [14] J.-P. Aubin and A. Cellina, *Differential inclusions: set-valued maps and viability theory*. Springer Science & Business Media, 2012, vol. 264.
- [15] S. Li, Y. Ogura, and V. Kreinovich, *Limit theorems and applications of set-valued and fuzzy set-valued random variables*. Springer Science & Business Media, 2013, vol. 43.
- [16] J.-P. Aubin, *Viability theory*. Springer Science & Business Media, 2009.
- [17] V. S. Borkar, “Optimal control of diffusion processes,” in *Pitman Research Notes in Math., 203. 36 Borkar V.(2005):“Controlled diffusion processes”, Probability surveys*. Citeseer, 1989.
- [18] D. P. Bertsekas, *Convex optimization theory*. Athena Scientific Belmont, 2009.
- [19] S. Perkins, D. S. Leslie *et al.*, “Asynchronous stochastic approximation with differential inclusions,” *Stochastic Systems*, vol. 2, no. 2, pp. 409–446, 2012.
- [20] A. Nagurney and D. Zhang, *Projected dynamical systems and variational inequalities with applications*. Springer Science & Business Media, 2012, vol. 2.